IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Ashkenazi et al.
App. No.	:	09/903,925
Filed	:	July 11, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

_____
(Date)

_____

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.

2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan® PCR or the fluorescence *in situ* hybridization (FISH) assays -is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi
Avi Ashkenazi, Ph.D.

Date: 9/15/03

SV 455281 v1
9/12/03 3:06 PM (39780.7000)

CURRICULUM VITAE

Avi Ashkenazi

July 2003

Personal:

Date of birth:	29 November, 1956
Address:	1456 Tarrytown Street, San Mateo, CA 94402
Phone:	(650) 578-9199 (home); (650) 225-1853 (office)
Fax:	(650) 225-6443 (office)
Email:	aa@gene.com

Education:

1983:	B.S. in Biochemistry, with honors, Hebrew University, Israel
1986:	Ph.D. in Biochemistry, Hebrew University, Israel

Employment:

1983-1986:	Teaching assistant, undergraduate level course in Biochemistry
1985-1986:	Teaching assistant, graduate level course on Signal Transduction
1986 - 1988:	Postdoctoral fellow, Hormone Research Dept., UCSF, and Developmental Biology Dept., Genentech, Inc., with J. Ramachandran
1988 - 1989:	Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc., with D. Capon
1989 - 1993:	Scientist, Molecular Biology Dept., Genentech, Inc.
1994 -1996:	Senior Scientist, Molecular Oncology Dept., Genentech, Inc.
1996-1997:	Senior Scientist and Interim director, Molecular Oncology Dept., Genentech, Inc.
1997-1990:	Senior Scientist and preclinical project team leader, Genentech, Inc.
1999 -2002:	Staff Scientist in Molecular Oncology, Genentech, Inc.
2002-present:	Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

Awards:

1988:	First prize, The Boehringer Ingelheim Award

Editorial:

Editorial Board Member: Current Biology
Associate Editor, Clinical Cancer Research.
Associate Editor, Cancer Biology and Therapy.

Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* **34**, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* **118**, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* **50**, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* **14**, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* **210**, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotropic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* **121**, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* **236**, 600-605 (1987).
8. Peralta, E. Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* **6**, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* **238**, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* **37**, 119-129 (1988).

11. Peralta, E. Ashkenazi, A., Winslow, J. Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* **334**, 434-437 (1988).

12. Ashkenazi., A. Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* **56**, 487-493 (1989).

13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* **340**, 146-150 (1989).

14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* **245**, 743-745 (1989).

15. Ashkenazi., A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodefficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* **87**, 7150-7154 (1990).

16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* **29**, 9885-9891 (1990).

17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodefficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* **88**, 7056-7060 (1991).

18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari., I., Pennica, D., Goeddel., D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* **88**, 10535-10539 (1991).

19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel., R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine$_{1B}$ receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).

21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).

22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).

23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT$_{1B}$ receptors. *Nature* **360**, 161-163 (1992).

24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon γ by an interferon γ receptor immunoadhesin. *Immunology* 79, 594-599 (1993).

25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).

26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).

27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3+ effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).

28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by iterferon γ and correction by interferon γ receptor immunoadhesin. *Blood* **83**, 911-915 (1994).

29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., andAshkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkilation. *Bioconj. Chem.* **5**, 133-140 (1994).

31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).

32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).

33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).

34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT2 receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).

35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN-γ receptor β chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).

36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).

37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA.* **92**, 5401-5405 (1995).

38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal E. coli bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).

39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* **6**, 1669-1676 (1996).

41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF-κB and AP-1 transcription factors. *J. Biol. Chem.* **272**, 14029-14032 (1997).

42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* **277**, 818-821 (1997).

43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Balswin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* **7**, 1003-1006 (1997).

44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* **8**, 525-528 (1998).

45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* **427**, 124-128 (1998).

46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* **160**, 4098-4103 (1998).

47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**, 699-703 (1998).

48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* **162**, 5616-5623 (1999).

49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* **9**, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).

51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN-α receptor 2. *J. Immunol.* **163**, 766-773 (1999).

52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).

53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).

54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563–571 (1999).

55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).

56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).

57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).

58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., *Ashkenazi, A., and *Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* **10**, 785-788 (2000).

60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* **29**, 506-512 (2000).

61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D. Mounho, B., Hillan, K., Totpal, K. DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* **7**, 383-385 (2001).

62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Martsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* **166**, 4891-4898 (2001).

63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* **7**, 1362-1369 (2001).

64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., *Ashkenazi, A., and *Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* **2**, 632-637 (2001).

65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).

66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotitc Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).

67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault., Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

Review articles:

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology.* **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels.* (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzimology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook (www.apnet.com/cytokinereference/).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. Cytokine and Growth Factor Reviews 14, 337-348 (2003).

Book:

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

Talks:

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Halflife Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficiacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon-γ signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Franciso, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: anovel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philladelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philladelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14th Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Wiezmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Swizerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

Issued Patents:

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6, 046, 048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).

2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.

4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.

5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

SV 2031808 v1

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South SanFrancisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985	Assistant Professor, Department of Chemistry, Oberlin College, Oberlin, Ohio
1980-1984	Graduate Research Assistant, Department of Biochemistry, Michigan State University East Lansing, Michigan

PUBLICATIONS:

1. **Polakis, P G.** and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun**. 107, 937-943.

2. **Polakis, P.G**. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys**. 234, 341-352.

3. **Polakis, P. G**. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys**. 236, 328-337.

4. Uhing, R.J., **Polakis,P.G**. and Snyderman, R. 1987 Isolaton of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem**. 262, 15575-15579.

5. **Polakis, P.G**., Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem**. 263, 4969-4979.

6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. **Polakis, P.G.**, Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun**. 161, 276-283.

8. **Polakis, P. G.**, Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Comun.** 160, 25-32.

9. **Polakis, P.**, Weber,R.F., Nevins,B., Didsbury, J. Evans,T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem**. 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.G.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Charaterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K^+ Channels. **Cell.** 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.

14. Polakis, P.G. Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits $p21^{ras}$ Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein $p21^{rap1}$. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for $p21^{ras}$-dependent Inhibition of Muscarinic Atrial K^+ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P** . 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 $p60^{v\text{-}src}$ Can Associate with and Phosphorylate the $p21^{ras}$ GTPase Activating Protein. **Oncogene** 7,389-394.

22. Polakis P. and McCormick, F. 1992 Interactions Between $p21^{ras}$ Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., **Polakis,P.** and McCormick, F. 1992 Molecular coloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.

24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle $p34^{cdc2}$ Kinase in vitro. **J. Biol. Chem**. 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., **Polakis, P.**, McCormick, F., and Innis, M. A. 1992 Functional Interactions Between $p21^{rap1A}$ and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol**. 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy,L., Clark, R., McCormick, F. and **Polakis, P**. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell . Biol**. 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., **Polakis, P.**, McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem**. 267, 25709-25713.

28. Janoueix-Lerosey, I., **Polakis, P.**, Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun**. 189, 455-464.

29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the *ras*-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of $p21^{ras}$ with GAP, exchange factors, and its bological effector target. **J. Biol Chem**. 268, 9157-9160.

31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and **Polakis, P.** 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., **Polakis, P.**, McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet**. 66, 18-21.

33. Sato, K. Y., **Polakis, P.**, Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor acitvity of the K-rev gene in human tumor cell lines. **Cancer Res**. 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., **Polakis, P.** and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun**. 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and **Polakis, P.** 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res**. 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31

37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71

38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α-catenin, β-catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555

39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β-catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.

40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.

41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.

42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.

43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β-catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.**16, 4088-4094.

44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein witha rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.

45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.

46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.

47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635

48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in Xenopus embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β-catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P** 1997 Loss of β-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes. K., Waterman, M., and **Polakis, P.** 1997 Induction of a β-catenin-LEF-1 complex by wnt-1 and transforming mutants of β-catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β-catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β-catenin and glycogen synthase kinase 3β. **Current Biology 8,** 573-581.

56. Polakis, P. 1998 The oncogenic activation of β-catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfled, Florence Margottin, Richard Benarous and **Paul Polakis**. 1999 The F-box protein β–TrCP associates with phosphorylated β–catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β–catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P**, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, V. ginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β-catenin-LEF signaling. **J. Biol. Chem**. 274(23):16641-5.

61 Polakis P, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol**. 470, 23-32

62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene**. 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett**. 458, 247-51.

65. Peifer M, **Polakis P**. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.

66. **Polakis P.** 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.

67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.

68. Szeto , W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.

69. Rubinfeld B, Tice DA, **Polakis P**. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.

70. **Polakis P**. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.

71. Tice DA, Soloviev I, **Polakis P**. 2002 Activation of the Wnt Pathway Interferes withSerum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem**. 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith J, Schwall RH, Pennnica D, **Polakis P**. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem**. 277,14329-14335.

73. **Polakis, P**. 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol**. 12, R499.

74. Mao,W. , Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., **Polakis, P.** 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res**. 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, **Polakis P**, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics**, April 8

SECOND DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I am currently employed by Genentech, Inc. where my job title is Staff Scientist.

2. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As I stated in my previous Declaration dated May 7, 2004 (attached as Exhibit A), my laboratory has been employing a variety of techniques, including microarray analysis, to identify genes which are differentially expressed in human tumor tissue relative to normal human tissue. The primary purpose of this research is to identify proteins that are abundantly expressed on certain human tumor tissue(s) and that are either (i) not expressed, or (ii) expressed at detectably lower levels, on normal tissue(s).

4. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor tissue at significantly higher levels than in normal human tissue. To date, we have successfully generated antibodies that bind to 31 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human tumor tissue and normal tissue. We have then quantitatively compared the levels of mRNA and protein in both the tumor and normal tissues analyzed. The results of these analyses are attached herewith as Exhibit B. In Exhibit B, "+" means that the mRNA or protein was detectably overexpressed in the tumor tissue relative to normal tissue and "-" means that no detectable overexpression was observed in the tumor tissue relative to normal tissue.

5. As shown in Exhibit B, of the 31 genes identified as being detectably overexpressed in human tumor tissue as compared to normal human tissue at the mRNA level, 28 of them (i.e., greater than 90%) are also detectably overexpressed in human tumor tissue as compared to normal human tissue at the protein level. As such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4-5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor tissue relative to a normal tissue more often than not correlates to a similar increase in abundance of the encoded protein in the tumor tissue relative to the normal tissue. In fact, it remains a generally accepted working assumption in molecular biology that increased mRNA levels are more often than not predictive of elevated levels of the encoded protein. In fact, an entire industry focusing on the research and development of therapeutic antibodies to treat a variety of human diseases, such as cancer, operates on this working assumption.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 3-29-06

By: Paul Polakis

Paul Polakis, Ph.D.

EXHIBIT A

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).

2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.

4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.

5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South SanFrancisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985	Assistant Professor, Department of Chemistry, Oberlin College, Oberlin, Ohio
1980-1984	Graduate Research Assistant, Department of Biochemistry, Michigan State University East Lansing, Michigan

PUBLICATIONS:

1. Polakis, P G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.

2. **Polakis, P.G.** and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.

3. **Polakis, P. G.** and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.

4. Uhing, R.J., **Polakis,P.G.** and Snyderman, R. 1987 Isolaton of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.

5. **Polakis, P.G.**, Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.

6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. **Polakis, P.G.**, Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.

8. **Polakis, P. G.**, Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Comun.** 160, 25-32.

9. Polakis, P., Weber,R.F., Nevins,B., Didsbury, J. Evans,T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.G.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Charaterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K^+ Channels. **Cell.** 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.

14. **Polakis, P.G.** Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol. 11,** 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits $p21^{ras}$ Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein $p21^{rap1}$. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for $p21^{ras}$-dependent Inhibition of Muscarinic Atrial K^+ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P.** 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 $p60^{v\text{-}src}$ Can Associate with and Phosphorylate the $p21^{ras}$ GTPase Activating Protein. **Oncogene** 7,389-394.

22. **Polakis P.** and McCormick, F. 1992 Interactions Between $p21^{ras}$ Proteins and Their GTPase Activating Proteins. In Cancer Surveys (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., **Polakis,P.** and McCormick, F. 1992 Molecular coloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.

24. **Polakis, P.**, Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle $p34^{cdc2}$ Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., **Polakis, P.**, McCormick, F., and Innis, M. A. 1992 Functional Interactions Between $p21^{rap1A}$ and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy,L., Clark, R., McCormick, F. and **Polakis, P.** 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell . Biol.** 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., **Polakis, P.**, McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.

28. Janoueix-Lerosey, I., **Polakis, P.**, Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.

29. **Polakis, P.** 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the *ras*-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. **Polakis, P.** and McCormick, F. 1993 Structural requirements for the interaction of $p21^{ras}$ with GAP, exchange factors, and its bological effector target. **J. Biol Chem.** 268, 9157-9160.

31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and **Polakis, P.** 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., **Polakis, P.**, McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.

33. Sato, K. Y., **Polakis, P.**, Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor acitvity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., **Polakis, P.** and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and **Polakis, P.** 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31

37. **Polakis, P.** 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71

38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α-catenin, β-catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555

39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β-catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.

40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.

41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.

42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.

43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β-catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.**16, 4088-4094.

44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein witha rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.

45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.

46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.

47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635

48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in Xenopus embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β-catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. **Polakis, P.** The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta,** 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P** 1997 Loss of β-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes. K., Waterman, M., and **Polakis, P.** 1997 Induction of a β-catenin-LEF-1 complex by wnt-1 and transforming mutants of β-catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.,** 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.,** 1998 Down regulation of β-catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β-catenin and glycogen synthase kinase 3β. **Current Biology 8,** 573-581.

56. **Polakis, P.** 1998 The oncogenic activation of β-catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfled, Florence Margottin, Richard Benarous and **Paul Polakis.** 1999 The F-box protein β–TrCP associates with phosphorylated β–catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β–catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P,** Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β-catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.

61 Polakis P, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32

62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.

65. Peifer M, **Polakis P.** 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.

66. Polakis P. 2000 Wnt signaling and cancer. Genes Dev;14, 1837-1851.

67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.

68. Szeto , W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.

69. Rubinfeld B, Tice DA, **Polakis P.** 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.

---

70. Polakis P. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.

71. Tice DA, Soloviev I, **Polakis P.** 2002 Activation of the Wnt Pathway Interferes withSerum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem.** 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennnica D, **Polakis P.** 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.

73. **Polakis, P.** 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao,W. , Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., **Polakis, P.** 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, **Polakis P,** Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics,** April 8

EXHIBIT B

	tumor mRNA	tumor IHC
UNQ2525	+	+
UNQ2378	+	+
UNQ972	+	-
UNQ97671	+	+
UNQ2964	+	+
UNQ323	+	+
UNQ1655	+	+
UNQ2333	+	+
UNQ9638	+	+
UNQ8209	+	+
UNQ6507	+	+
UNQ8196	+	+
UNQ9109	+	+
UNQ100	+	+
UNQ178	+	+
UNQ1477	+	+
UNQ1839	+	+
UNQ2079	+	+
UNQ8782	+	+
UNQ9646	+	-
UNQ111	+	+
UNQ3079	+	+
UNQ8175	+	+
UNQ9509	+	+
UNQ10978	+	-
UNQ2103	+	+
UNQ1563	+	+
UNQ16188	+	+
UNQ13589	+	+
UNQ1078	+	+
UNQ879	+	+

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman||, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. ***Molecular & Cellular Proteomics 1:37–45, 2002.***

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1*, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)[1] have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q−, 11p−, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

From the ‡Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and **Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ¶AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ||UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

[1] The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

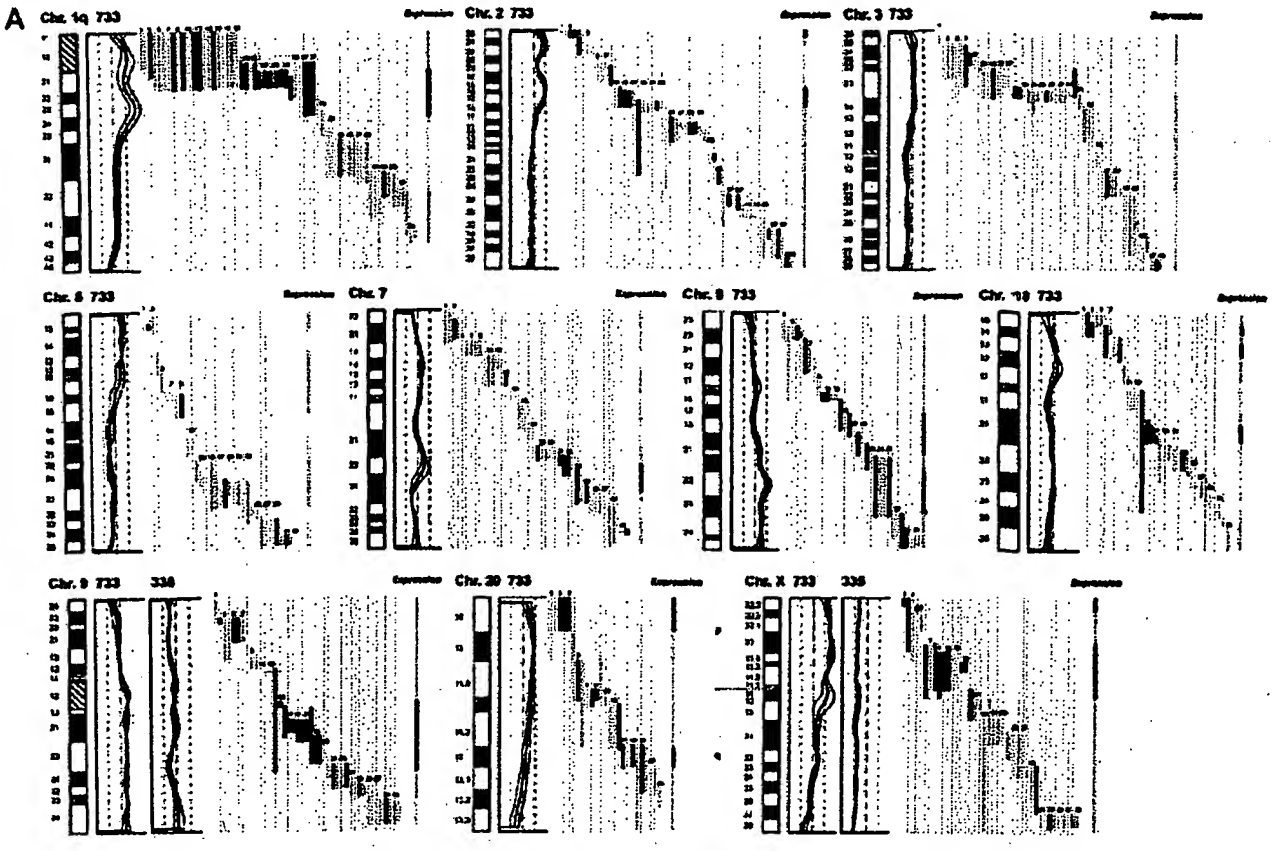

FIG. 1. **DNA copy number and mRNA expression level.** Shown from *left* to *right* are chromosome (*Chr.*), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. *A*, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. *B*, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (*left*). The *bold curve* in the ratio profile represents a mean of four chromosomes and is surrounded by *thin curves* indicating one standard deviation. The *central vertical line* (*broken*) indicates a ratio value of 1 (no change), and the *vertical lines* next to it (*dotted*) indicate a ratio of 0.5 (*left*) and 2.0 (*right*). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the *right* of the invasive tumor profile. The *colored bars* represents one gene each, identified by the running *numbers above* the *bars* (the name of the gene can be seen at www.MDL.DK/sdata.html). The *bars* indicate the purported location of the gene, and the *colors* indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (*black*), >2-fold decrease (*blue*), no significant change (*orange*). The *bar* to the *far right*, entitled *Expression* shows the resulting change in expression along the chromosome; the *colors* indicate that at least half of the genes were up-regulated (*black*), at least half of the genes down-regulated (*blue*), or more than half of the genes are unchanged (*orange*). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at −80 °C. Total RNA was isolated using the RNAzol B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)$^+$ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript® choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the MEGAscrip® *in vitro* transcription kit (Ambion). Biotin-labeled CTP and UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94 °C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6× SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95 °C for 5 min, subsequently cooled to 40 °C, and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40 °C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6× SSPE-T at 25 °C followed by 4 washes in 0.5× SSPE-T at 50 °C. The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 μg/ml (Molecular Probes) in 6× SSPE-T

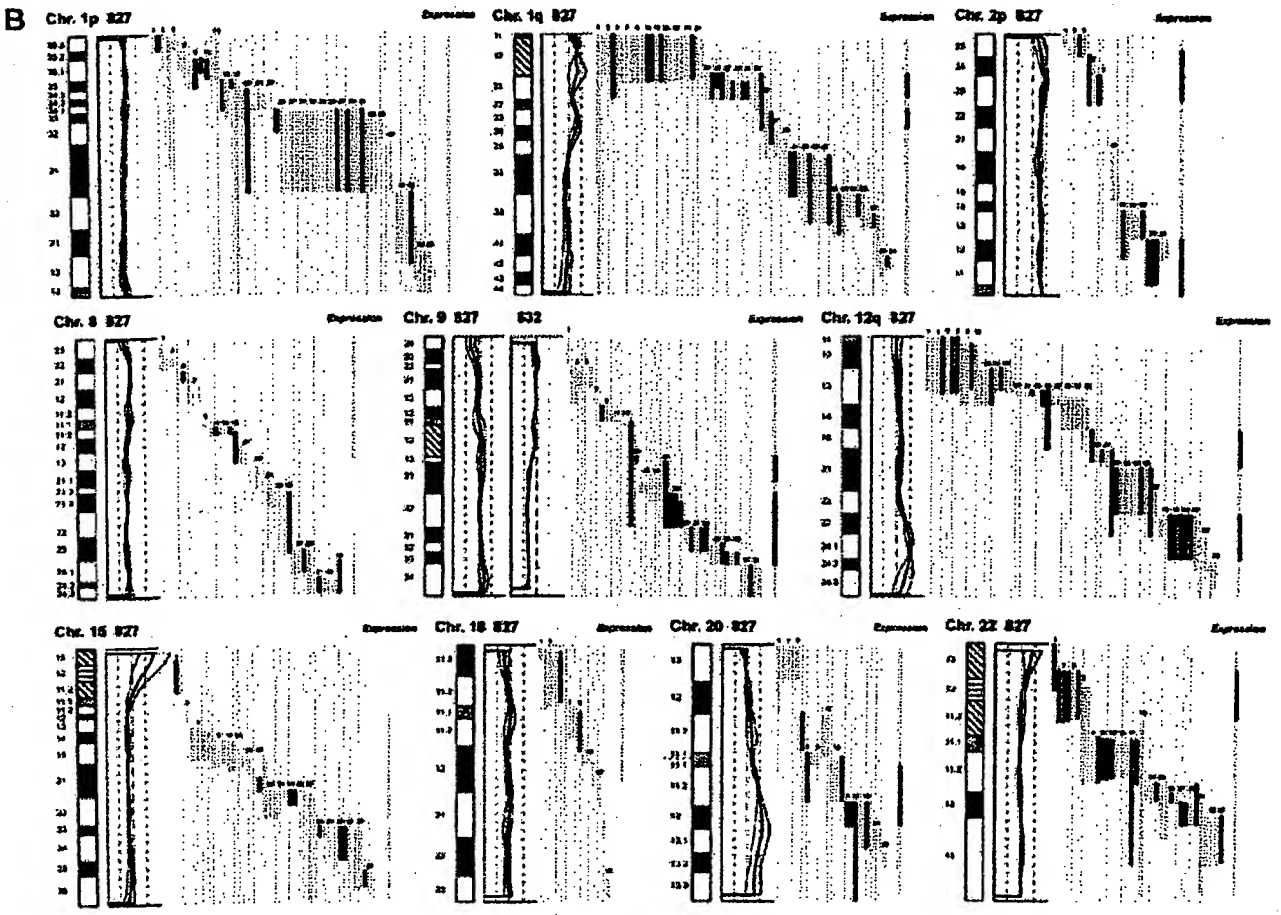

FIG. 1—*continued*

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μl for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at −20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/cells.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-labeled reference DNA (200 ng), and human Cot-1 DNA (20 μg) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μg/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

TABLE I

Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%

Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22-q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1-qter–, 3q12-q13.3–, 6q12-q22–, 9q34+, 11q12-q13+, 17+, and 20q11.2-q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1*A*) and 20q12 in TCC 827 (Fig. 1*B*).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21-q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, *top*). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, *top*). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, *bottom*). Furthermore, as a control we looked at areas with no alter-

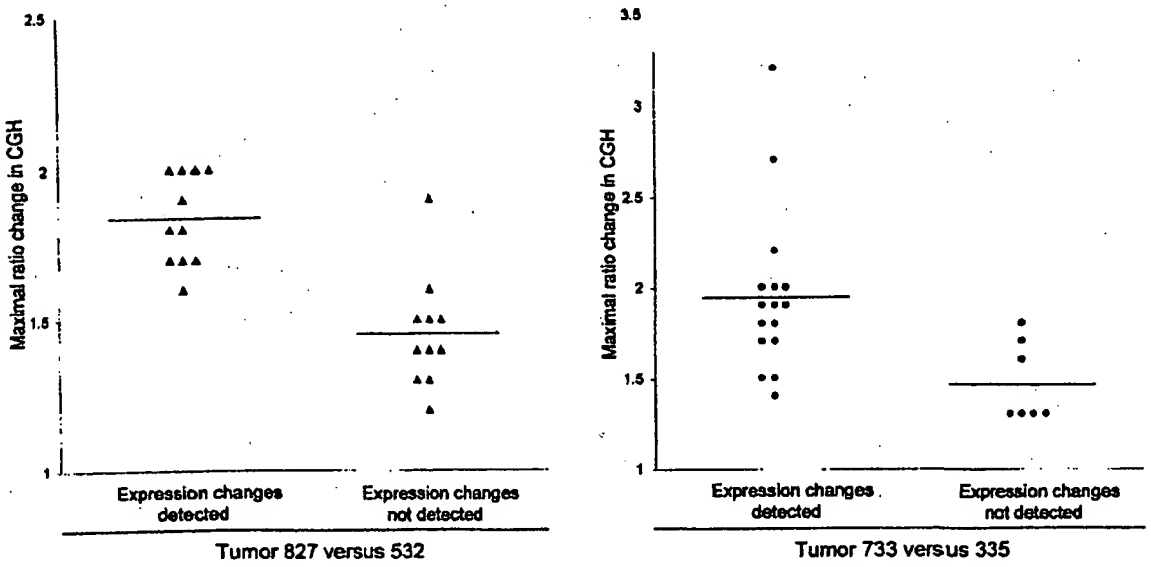

FIG. 2. **Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring.** The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (♦) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the *right* in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1*A*). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

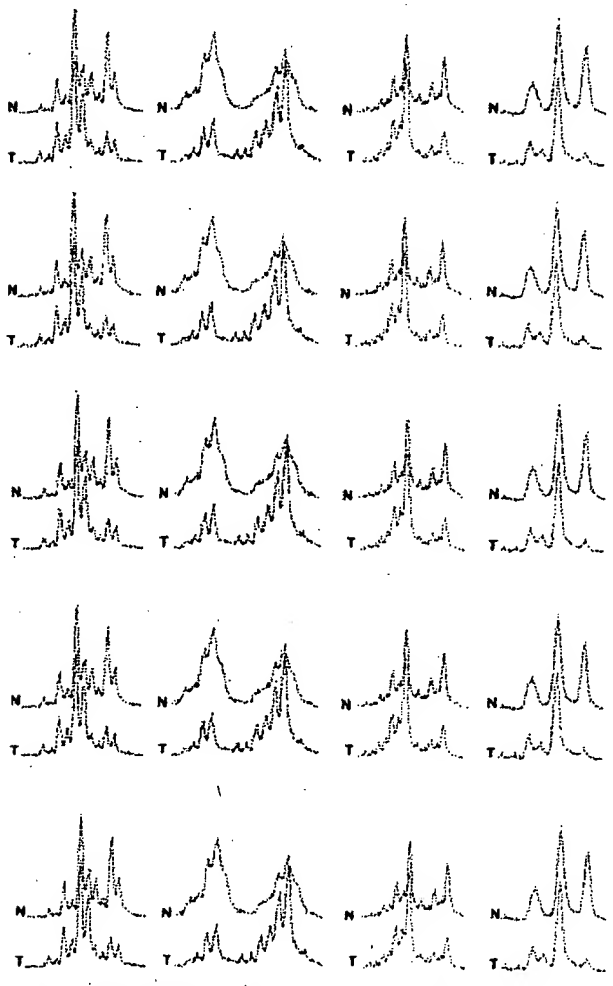

FIG. 3. **Microsatellite analysis of loss of heterozygosity.** Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (*a*) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (*b*) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (*c*) at chromosome 2p23 by D2S2251 close to general β-spectrin (gene number 11 on Fig. 1) and of (*d*) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The *upper curves* show the electropherogram obtained from normal DNA from leukocytes (*N*), and the *lower curves* show the electropherogram from tumor DNA (*T*). In all cases one allele is partially lost in the tumor amplicon.

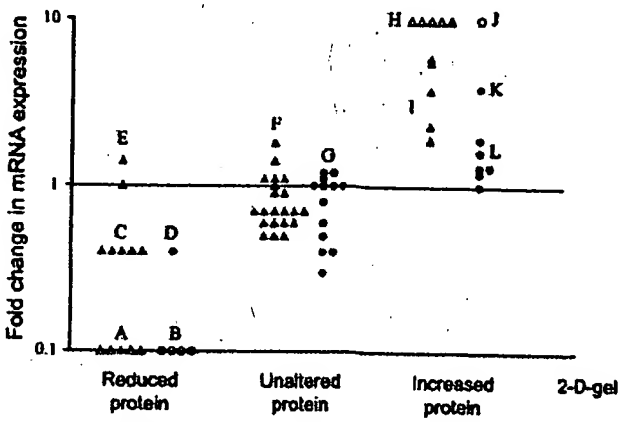

FIG. 4. **Correlation between protein levels as judged by 2D-PAGE and transcript ratio.** For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (*horizontal axis*). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (*vertical axis*). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; △, mRNAs that were scored as absent in the invasive tumors (along *horizontal axis*) or as absent in non-invasive reference (*top* of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲△) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group *A* (from *left*), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytokeratin 15, and cytokeratin 13; *B* (from *left*), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytokeratin 13, and calcyclin; *C* (from *left*), α-enolase, hnRNP B1, 28-kDa heat shock protein, 14–3-3-ε, and pre-mRNA splicing factor; *D*, mesothelial keratin K7 (type II); *E* (from *top*), glutathione *S*-transferase-π and mesothelial keratin K7 (type II); *F* (from *top* and *left*), adenylyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ-actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β-1 subunit; *G*, (from *top* and *left*), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytokeratin 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione *S*-transferase-π, and keratin 8; *H* (from *left*), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15 hydroxyprostaglandin dehydrogenase; *I* (from *top*), prolyl 4-hydroxylase β-subunit, cytokeratin 20, cytokeratin 17, prohibition, and fructose 1,6-biphosphatase; *J* annexin II; *K*, annexin IV; *L* (from *top* and *left*), 90-kDa heat shock protein, prolyl 4-hydroxylase β-subunit, α-enolase, GRP 78, cyclophilin, and cofilin.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

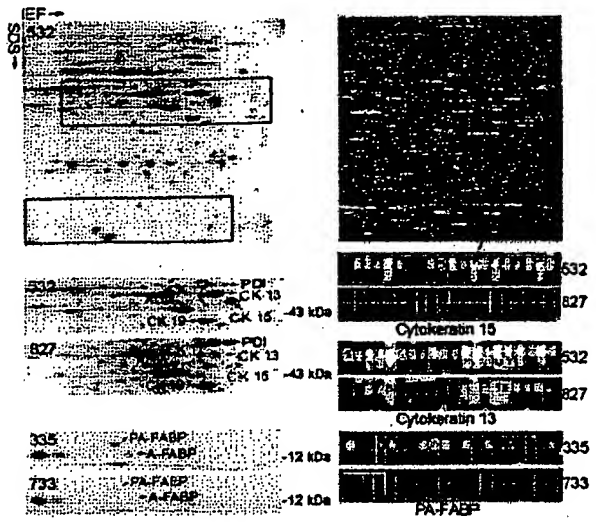

FIG. 5. **Comparison of protein and transcript levels in invasive and non-invasive TCCs.** The *upper part* of the figure shows a 2D gel (*left*) and the oligonucleotide array (*right*) of TCC 532. The *red rectangles* on the *upper gel* highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (*red annotation*). The tile on the array containing probes for cytokeratin 15 is enlarged *below* the array (*red arrow*) from TCC 532 and is compared with TCC 827. The *upper row* of *squares* in each tile corresponds to perfect match probes; the *lower row* corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as *black*, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the *bottom* of the figure (*left*) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the *right* we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). *IEF*, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a know chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II

Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration[a]	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres[a]	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up[b]	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

[a] Abs, absent; Pres, present.

[b] In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (*e.g.* in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p−, 9q−, 1q+, Y− (2, 6), and in pT1 tumors, 2q−,11p−, 11q−, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p− and 9q22-q33− and 9q− and Y−, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p−, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicate that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (*e.g.* signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Vikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

REFERENCES

1. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* **17**, 643–649
2. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46
3. de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Llva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* **83**, 157–161
4. Brungier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* **12**, 1747–1753
5. Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* **50**, 1459–1463
6. Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) c-*myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am J. Pathol.* **146**, 1131–1139
7. Richter, J., Jiang, F., Gorog, J. P., Sartorious, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* **57**, 2860–2864
8. Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* **153**, 1615–1621
9. Bruch, J., Wohr, G., Hautmann, R., Matthaeidt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* **23**, 167–174
10. Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* **15**, 3555–3560
11. Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* **185**, 345–351
12. Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* **110**, 87–93
13. Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367
14. Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* **59**, 167–177
15. Celis, J. E., Ostergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* **56**, 4782–4790
16. Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) In *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL
17. Ohlsson, R., Tycko, B., and Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* **14**, 435–438
18. Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121
19. Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170
20. Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpa, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* **59**, 3889–3892
21. Cui, H., Hedborg, F., He, L., Nordenskjold, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* **57**, 4469–4473
22. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* **285**, 251–254
23. Tsao, J., Yatabe, Y., Markl, I. D., Hajyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* **29**, 26–32
24. Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10632–10636
25. Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
26. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934

[CANCER RESEARCH 62, 6240–6245, November 1, 2002]

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer[1,2]

Elizabeth Hyman,[3] Päivikki Kauraniemi,[3] Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi[4]

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E. H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E. H., P. K., S. H., M. W., S. M., E. R., M. R., A. E., O. K., A. K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P. K., A. K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S. H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G. S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O. M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1–6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over 20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH[5] (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (*a*) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (*b*) identify and characterize those genes whose mRNA expres-

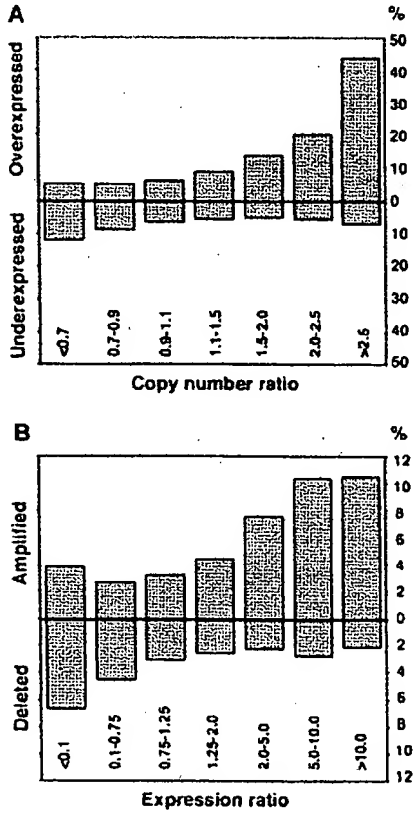

Fig. 1. Impact of gene copy number on global gene expression levels. *A*, percentage of over- and underexpressed genes (*Y axis*) according to copy number ratios (*X axis*). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). *B*, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7.

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

[2] Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org).

[3] Contributed equally to this work.

[4] To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkkeilijankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

[5] The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription-PCR.

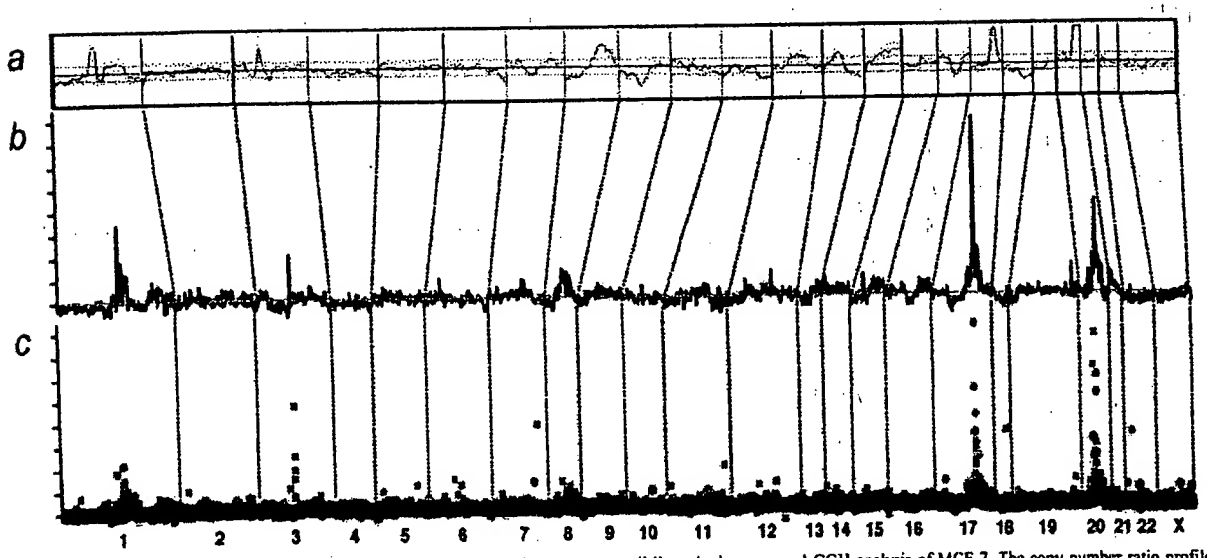

Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (*blue line*) across the entire genome from 1p telomere to Xq telomere is shown along with ±1 SD (*orange lines*). The *black horizontal line* indicates a ratio of 1.0; *red line*, a ratio of 0.8; and *green line*, a ratio of 1.2. *B–C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. *Red horizontal line*, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The *bright red dots* indicate the upper 2%, and *dark red dots*, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); *bright green dots* indicate the lowest 2%, and *dark green dots*, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with *black crosses*. The chromosome numbers are shown at the *bottom* of the figure, and chromosome boundaries are indicated with a *dashed line*.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μg of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *Alu*I and *Rsa*I (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μg of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μg of reference RNA were labeled with Cy3-dUTP and 3.5 μg of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units) were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g, for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1}, σ_{g1} and m_{g0}, σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α. A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.[6] A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.[7] The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

[6] Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.
[7] Internet address: www.genome.ucsc.edu.

Table 1 *Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray*

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1–7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11–8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3–8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22–q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12–q21.2	39.79	42.80	3.0
17q21.32–q21.33	52.47	55.80	3.3
17q22–q23.3	63.81	69.70	5.9
17q23.3–q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12–q13.13	46.45	49.43	3.0
20q13.2–q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5′-GAGCAGAGGGACTCGGACTT-3′ and 5′-GCGTCAGGTAGCGATTGTAG-3′.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (*i.e.*, belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1*A*). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1*B*). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal CGH were validated, with 1q21, 17q12—q21.2, 17q22–q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65–39.25 Mb), and 17q21.3 (52.47–55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2*C* shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22–q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11–p12 (Fig. 3*A*). In BT-474, the two known amplicons at 17q12 and 17q22–q23 contained numerous highly overexpressed genes (Fig. 3*B*). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3*B*, *inset*) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

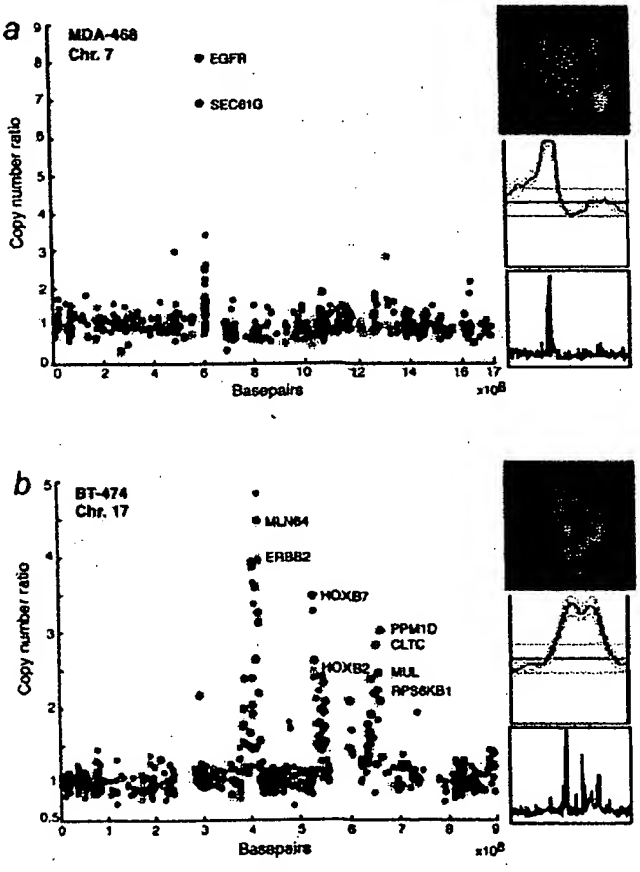

Fig. 3. Annotation of gene expression data on CGH microarray profiles. *A*, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (*red dots*) and include the *EGFR* oncogene. *B*, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (*red*) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2*C*. *Insets* show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (*red*) and chromosome 7 centromere probe (*green*) to MDA-468 (*A*) and *HOXB7*-specific probe (*red*) and chromosome 17 centromere (*green*) to BT-474 cells (*B*).

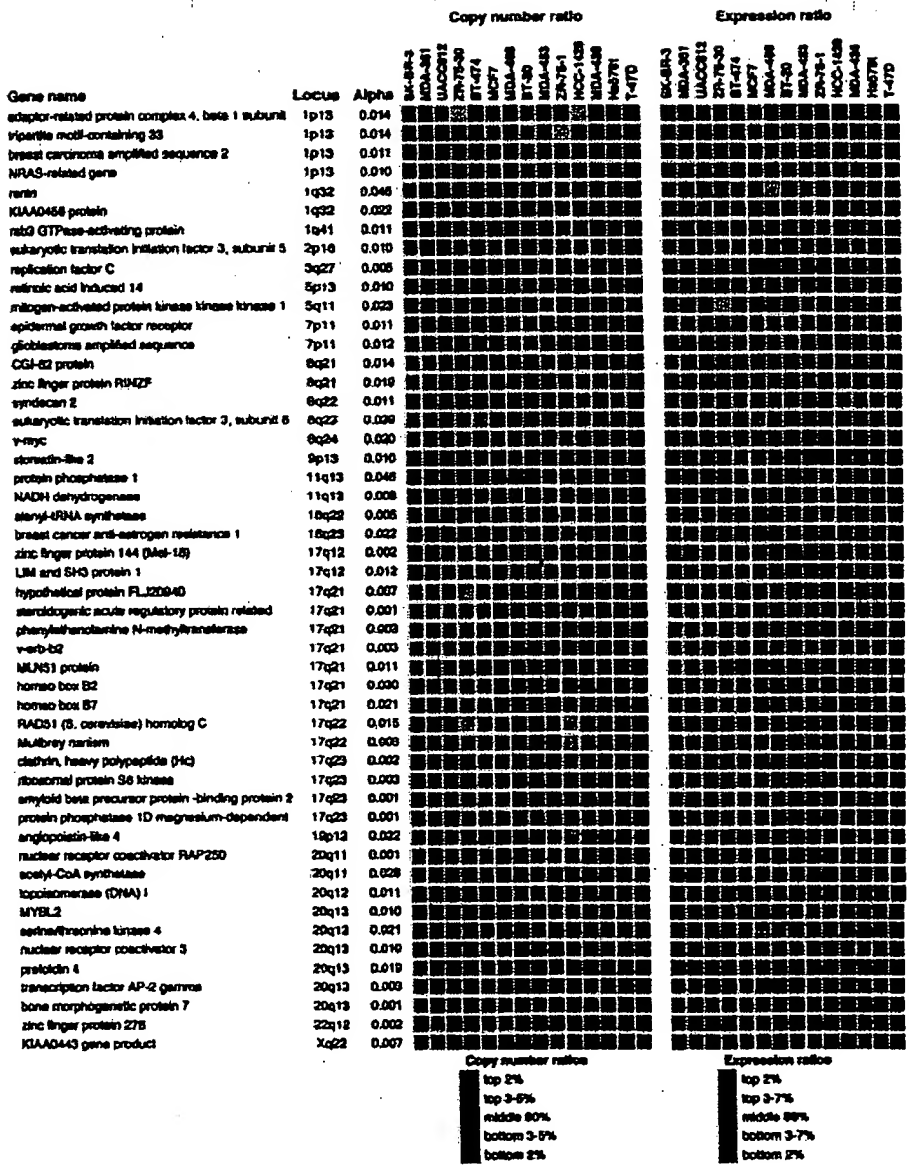

Fig. 4. List of 50 genes with a statistically significant correlation (α value <0.05) between gene copy number and gene expression. Name, chromosomal location, and the α value for each gene are indicated. The genes have been ordered according to their position in the genome. The color maps on the *right* illustrate the copy number and expression ratio patterns in the 14 cell lines. The key to the color code is shown at the *bottom* of the graph. *Gray squares*, missing values. The complete list of 270 genes is shown in supplemental Fig. B.

amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ($P = 0.001$).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,[8] 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH[9] (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

[8] Internet address: http://www.geneontology.org/.

[9] Internet address: http://www.ncbi.nlm.nih.gov/entrez.

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein s6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (*a*) evidence of a prominent global influence of copy number changes on gene expression levels; (*b*) a high-resolution map of 24 independent amplicons in breast cancer; and (*c*) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

1. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.
2. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), *403:* 503–511, 2000.
3. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., *et al.* Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature (Lond.), *406:* 536–540, 2000.
4. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.
5. Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. Nature (Lond.), *412:* 822–826, 2001.
6. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA, *98:* 10869–10874, 2001.
7. Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. Semin. Cancer Biol., *9:* 125–138, 1999.
8. Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. J. Clin. Oncol., *19:* 32–40, 2001.
9. Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., *et al.* DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am. J. Pathol., *152:* 1107–1123, 1998.
10. Knuutila S., Autio K., and Aalto Y. Online access to CGH data of DNA sequence copy number changes. Am. J. Pathol., *157:* 689, 2000.
11. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.
12. Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res., *6:* 639–645, 1996.
13. Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. *In:* F. J. Livesey and S. P. Hunt (eds.), Functional Genomics. pp. 113–137. Oxford: Oxford University Press, 2000.
14. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet., *23:* 41–46, 1999.
15. Monni, O., Bärlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc. Natl. Acad. Sci. USA, *98:* 5711–5716, 2001.
16. Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics, *2:* 364–374, 1997.
17. Bärlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., *et al.* Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J. Natl. Cancer Inst., *92:* 1252–1259, 2000.
18. Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. Cytometry, *45:* 83–86, 2001.
19. Kauraniemi, P., Bärlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res., *61:* 8235–8240, 2001.
20. Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer, *34:* 104–114, 2002.
21. Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res., *62:* 2625–2629, 2002.
22. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet., *25:* 333–337, 2000.
23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc. Natl. Acad. Sci. USA, *98:* 1124–1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linehan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res., *61:* 8143–8149, 2001.
25. Bärlund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22–q24 by CGH in breast cancer is due to high-level amplification of two separate regions. Genes Chromosomes Cancer, *20:* 372–376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. Cancer Res., *56:* 3441–3445, 1996.
27. Cillo, C., Faicila, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. Exp. Cell Res., *248:* 1–9, 1999.
28. Cillo, C., Cantile, M., Faicila, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. J. Cell. Physiol., *188:* 161–169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. Mol. Cell. Biol., *16:* 4842–4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the *HOXB7* gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. Oncogene, *16:* 3285–3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. Cancer Res., *61:* 6532–6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. Proc. Natl. Acad. Sci. USA, *98:* 4060–4065, 2001.

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack*†‡, Therese Sørlie§, Charles M. Perou¶, Christian A. Rees‖**, Stefanie S. Jeffrey††, Per E. Lonning‡‡, Robert Tibshirani§§, David Botstein‖, Anne-Lise Børresen-Dale§, and Patrick O. Brown†¶¶

Departments of *Pathology, ‖Genetics, ††Surgery, §§Health Research and Policy, and ¶¶Biochemistry, and †Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; §Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ‡‡Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ¶Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. "Test" DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The "reference" (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at http://rana.lbl.gov). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

*To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

**Present address: Zyomyx Inc., Hayward, CA 94545.

GENETICS

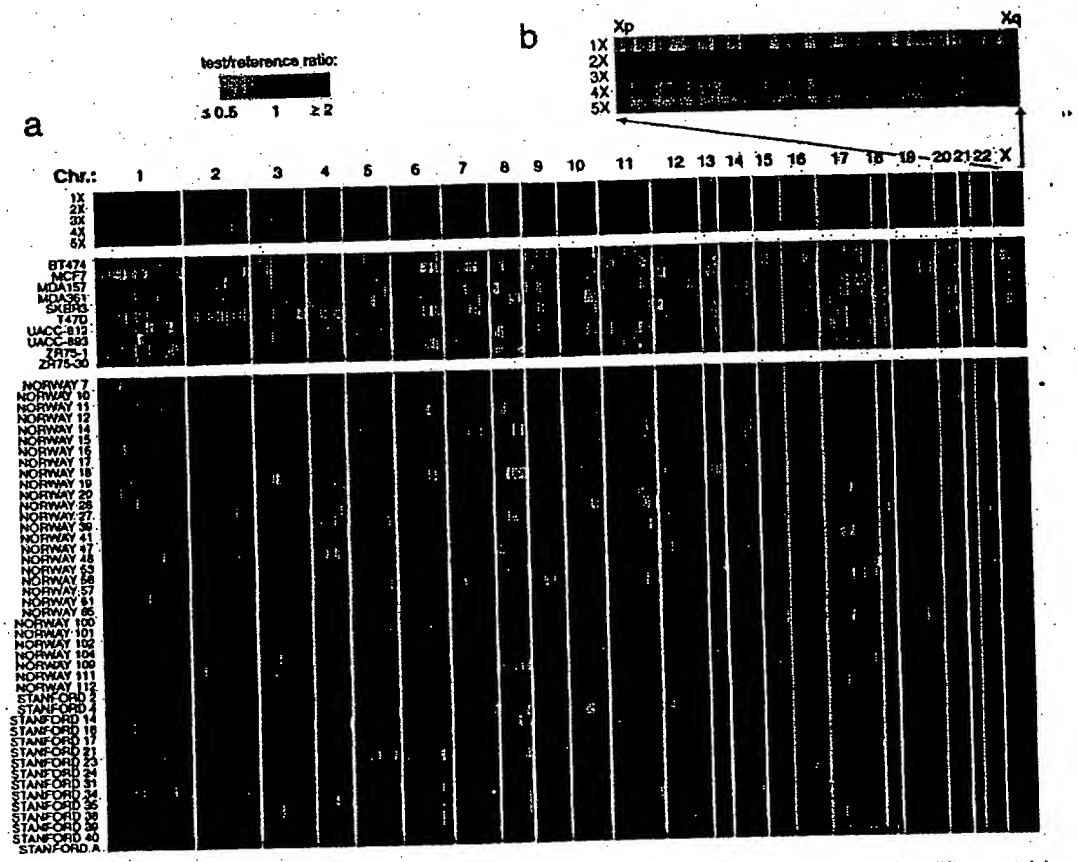

Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (*a*) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a $\log_2$-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (*b*) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (http://genome.ucsc.edu/; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1*a*; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (http://genome.ucsc.edu/) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1*b*), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1*a*), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2–4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total

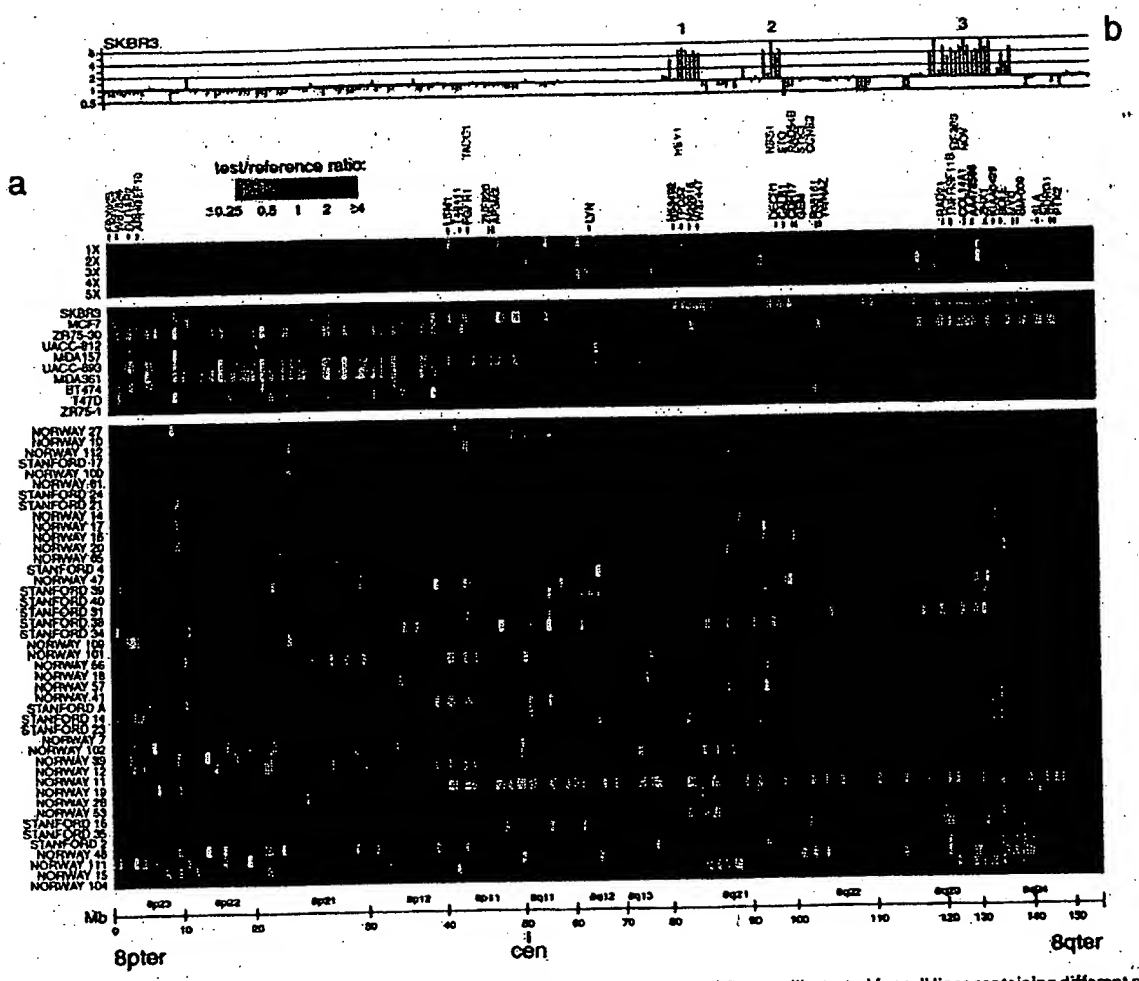

Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (*a*) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a $\log_2$ pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (*b*) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a $\log_2$ scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2*a*). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2*b*). For each of these regions we can define the boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

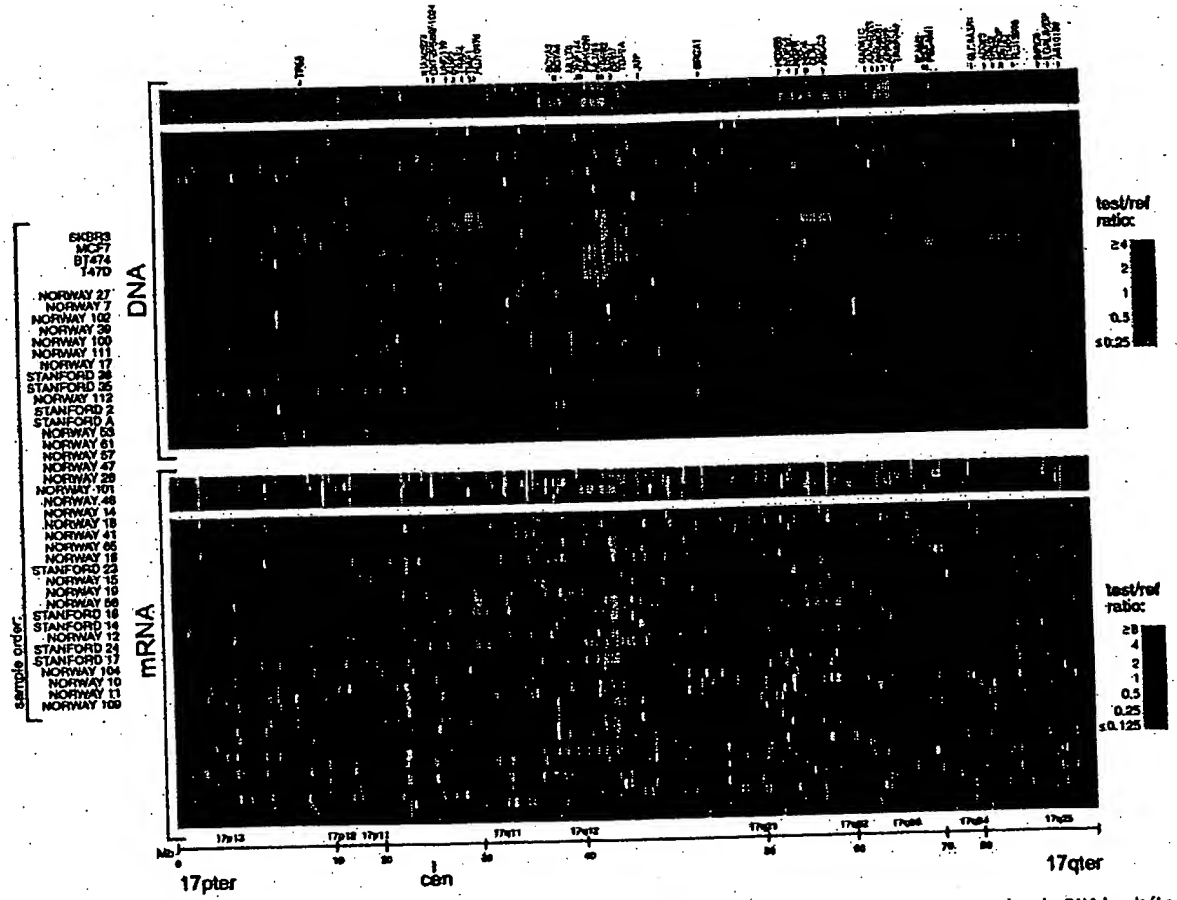

Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (*Upper*) and mRNA levels (*Lower*) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (*Upper*), and the identical sample order is maintained (*Lower*). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate $\log_2$ pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4, and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4*a*). For both the breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (P values for pair-wise Student's t tests comparing adjacent classes: cell lines, 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2}; tumors, 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4}). A linear regression of the average log(DNA copy number), for each class, against average log(mRNA level) demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4*a*, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4*b*). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4*c*), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4*b*) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37

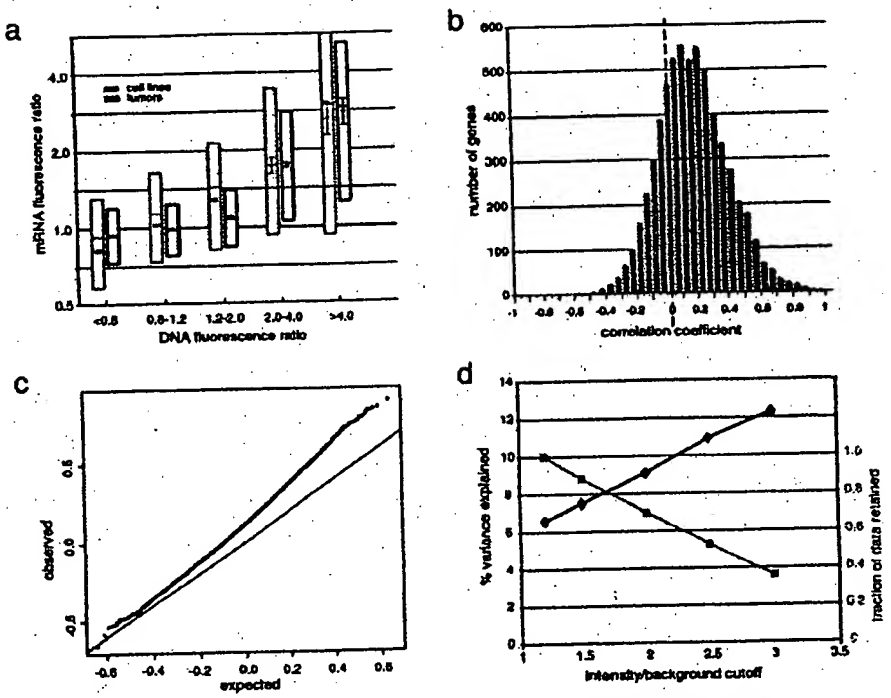

Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (*a*) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log_2 scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2} (cell lines), and 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4} (tumors). (*b*) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (*c*) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (*d*) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4*d*). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background >3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4*d*). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stochiometric relationships in cell metabolism and physiology (e.g., proteosome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818–821.
2. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156–2160.
3. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177–184.
4. Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519–4525.
5. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 20, 399–407.
6. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207–211.
7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41–46.
8. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747–752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869–10874.
10. Schuler, G. D. (1997) *J. Mol. Med.* 75, 694–698.
11. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860–921.
12. Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105–113.
13. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333–337.
14. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143–8149.
15. Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134–1138.
16. Albertson, D. G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144–146.
17. Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506–14511.
18. Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 340, 621–630.

TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

HER-2/neu Breast Cancer Predictive Testing

Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.[1] Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.[2]

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.[3]

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.[4] HER-2/neu status may be particularly important to establish in women with small ($\leq$1 cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.[5] Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin© (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin© in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin© in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest©) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

CPT code information

HER-2/neu via IHC

88342 (including interpretive report)

HER-2/neu via FISH

88271×2 Molecular cytogenetics, DNA probe, each
88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells
88291 Cytogenetics and molecular cytogenetics, interpretation and report

Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest©. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion™ HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The Pathvysion™ kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2 "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3 Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4 Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-s/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5 Press, M.F. Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

Provided for the clients of

PATHOLOGY ASSOCIATES MEDICAL LABORATORIES
PACLAB NETWORK LABORATORIES
TRI-CITIES LABORATORY
TREASURE VALLEY LABORATORY

For more information, please contact your local representative.

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4049–4052, June 1986
Medical Sciences

Variable expression of the translocated c-*abl* oncogene in Philadelphia-chromosome-positive B-lymphoid cell lines from chronic myelogenous leukemia patients

JAMES B. KONOPKA*‡, STEVEN CLARK*, JAMI MCLAUGHLIN*, MASAKUZU NITTA†, YOSHIRO KATO†, ANNABEL STRIFE†, BAYARD CLARKSON†, AND OWEN N. WITTE*§

*Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024; and †The Laboratory of Hematopoietic Cell Kinetics and The Laboratory of Cancer Genetics and Cytogenetics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

Communicated by Michael Potter, February 10, 1986

ABSTRACT **The consistent cytogenetic translocation of chronic myelogenous leukemia (the Philadelphia chromosome, Ph[1]) has been observed in cells of multiple hematopoietic lineages. This translocation creates a chimeric gene composed of breakpoint-cluster-region (*bcr*) sequences from chromosome 22 fused to a portion of the *abl* oncogene on chromosome 9. The resulting gene product (P210$^{c\text{-}abl}$) resembles the transforming protein of the Abelson murine leukemia virus in its structure and tyrosine kinase activity. P210$^{c\text{-}abl}$ is expressed in Ph[1]-positive cell lines of myeloid lineage and in clinical specimens with myeloid predominance. We show here that Epstein–Barr virus-transformed B-lymphocyte lines that retain Ph[1] can express P210$^{c\text{-}abl}$. The level of expression in these B-cell lines is generally lower and more variable than that observed for myeloid lines. Protein expression is not related to amplification of the *abl* gene but to variation in the level of *bcr-abl* mRNA produced from a single Ph[1] template.**

Chronic myelogenous leukemia (CML) is a disease of the pluripotent stem cell (1). In greater than 95% of patients, the leukemic cells contain the cytogenetic marker known as the Philadelphia chromosome, or Ph[1] (2). This reciprocal translocation event between the long arms of chromosomes 9 and 22 has been used as a disease-specific marker for diagnosis and evaluation of therapy. Multiple hematopoietic lineages, including myeloid and B-lymphoid, contain Ph[1] in early or chronic phase, as well as in the more acute accelerated and blast crisis phases of the disease.

One molecular consequence of Ph[1] is the translocation of the chromosomal arm containing the c-*abl* gene on chromosome 9 into the middle of the breakpoint-cluster region (*bcr*) gene on chromosome 22 (3–6). Although the precise translocation breakpoints are variable, an RNA-splicing mechanism generates a very similar 8-kilobase (kb) mRNA in each case (5–9). The hybrid *bcr-abl* message encodes a structurally altered form of the *abl* oncogene product, called P210$^{c\text{-}abl}$ (10–13), with an amino-terminal segment derived from a portion of the exons of *bcr* on chromosome 22 and a carboxyl-terminal segment derived from a major portion of the exons of the c-*abl* gene on chromosome 9. The chimeric structure of *bcr-abl* and the resulting P210$^{c\text{-}abl}$ is similar to the structure of the Abelson murine leukemia virus *gag-abl* genome and resulting P160$^{v\text{-}abl}$ transforming gene product. Both proteins have very similar tyrosine kinase activities (10, 11, 14) which can be distinguished by their relative stability to denaturing detergents and by their ATP requirements from the recently described tyrosine kinase activity of the c-*abl* gene product (15).

In concert with structural modification of the amino-terminal portion of the *abl* gene, increased level of expression has been implicated in activation of c-*abl* oncogenic potential. Myeloid and erythroid cell lines and clinical samples derived from acute-phase CML patients contain about 10-fold higher levels of the 8-kb *bcr-abl* mRNA and P210$^{c\text{-}abl}$ than the c-*abl* mRNA forms (6 and 7 kb) and P145$^{c\text{-}abl}$ gene product (5, 8, 9, 11). The higher level of expression of the chimeric *bcr-abl* message in acute-phase cells is not likely to be solely due to the presence of the *bcr* promoter sequences at the 5′ end of the gene, since the normal 4.5-kb and 6.7-kb *bcr*-encoded mRNA species are expressed at an even lower level than the normal c-*abl* messages (5, 6).

We have analyzed a series of Epstein–Barr virus-immortalized B-lymphoid cell lines derived from CML patients (16). With such *in vitro* clonal cell lines, we can evaluate whether the presence of Ph[1] always results in synthesis of the chimeric *bcr-abl* message and protein, and whether the quantitative expression varies for cells of B-lymphoid lineage as compared to previously examined myeloid cell lines. Our results show that cell lines that retain Ph[1] do express *bcr-abl* message and protein, but that the level is generally lower and more variable than previously seen for myeloid cell lines. The demonstration that the Ph[1] chromosomal template can vary in its level of expression of P210$^{c\text{-}abl}$ suggests that secondary mechanisms, beyond the translocation itself, contribute to the regulation of the *bcr-abl* gene in different cell types or subclones that derive from the affected stem cell.

MATERIALS AND METHODS

Cells and Cell Labelings. Epstein–Barr virus-transformed B-lymphoid cell lines were established from peripheral blood samples of chronic- and acute-phase CML patients as reported (16). The cell lines are designated according to patient number, karyotype, and lineage. For example, SK-CML7Bt(9,22)-33 refers to CML patient 7, B-lymphoid cell line, 9;22 translocation (Ph[1]), cell line 33; and SK-CML7BN-2 refers to B-cell line 2 with a normal karyotype derived from the same patient. Repeat karyotype analysis was performed to verify the retention of Ph[1] just prior to analysis for *abl* protein and RNA. Cells were maintained in RPMI 1640 medium with 20% fetal bovine serum. We have not observed any consistent pattern of *in vitro* growth rate that correlates to the stage of disease at the time of transformation with Epstein–Barr virus. Cells (1.5×10^7) were washed twice with Dulbecco's modified Eagle's medium lacking phosphate and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *bcr*, breakpoint-cluster region; CML, chronic myelogenous leukemia; kb, kilobase(s).

‡Present address: Department of Genetics, University of Washington, Seattle, WA 98195.
§To whom correspondence should be addressed.

supplemented with 5% dialyzed fetal bovine serum. Cells were then resuspended in 2 ml of the minimal medium. Labeling was started with the addition of [^{32}P]orthophosphate (1 mCi/ml; ICN; 1 Ci = 37 GBq) and continued at 37°C for 3–4 hr.

Immunoprecipitation and Immunoblotting. Immunoprecipitations were carried out as described (10). Cells (1.5×10^7) were washed with phosphate-buffered saline and extracted with 3–5 ml of phosphate lysis buffer (1% Triton X-100/0.1 $NaDodSO_4$/0.5% deoxycholate/10 mM Na_2HPO_4, pH 7.5/100 mM NaCl) with 5 mM EDTA and 5 mM phenylmethylsulfonyl fluoride. Extracts were clarified by centrifugation and precipitated with normal or rabbit anti-abl sera (anti-pEX-2 or anti-pEX-5) (17). The precipitated proteins were electrophoresed in a $NaDodSO_4$/8% polyacrylamide gel. ^{32}P-labeled proteins were detected by autoradiography. Alternatively, *abl* proteins were detected by immunoblotting. Extracts from unlabeled cells were clarified, and proteins were concentrated by immunoprecipitation with rabbit antisera against *abl*-encoded proteins [anti-pEX-2 and anti-pEX-5 combined (17)] and then fractionated in 8% acrylamide gels. The proteins were transferred from the gel to nitrocellulose filters, using protease-facilitated transfer (18). The *abl*-encoded proteins were detected using murine monoclonal antibodies as a probe and peroxidase-conjugated goat anti-mouse second stage antibody (Bio-Rad) for development. Rabbit antisera and mouse monoclonal antibodies to *abl* proteins were prepared using bacterially expressed regions of the v-*abl* protein as immunogens (17, 19). Anti-pEX-2 antibodies react with the internal tyrosine kinase domain and anti-pEX-5 antibodies react with the carboxyl-terminal segment of the *abl* proteins.

RNA Analysis. RNA was extracted from 10^8 cells by the $NaDodSO_4$/urea/phenol method (20). Polyadenylylated RNA was purified by oligo(dT) affinity chromatography. Samples were electrophoresed in a 1% agarose/formaldehyde gel and transferred to nitrocellulose. *abl* RNA species were detected by hybridization with a nick-translated v-*abl* fragment probe (21).

DNA Analysis. DNA was prepared from 5×10^7 cells of each cell line and processed for Southern blots with a v-*abl* probe as described (21).

RESULTS

Variable Levels of P210$^{c\text{-}abl}$ Are Detected in Ph^1-Positive Cell Lines. Ph^1-positive and Ph^1-negative, Epstein–Barr virus-transformed B-lymphocyte cell lines derived from the same patient were examined for P210$^{c\text{-}abl}$ synthesis by immunoprecipitation of [^{32}P]orthophosphate-labeled cell extracts with anti-abl sera (Fig. 1). The normal c-*abl* protein P145$^{c\text{-}abl}$ was detected at a similar level in multiple Ph^1-positive and Ph^1-negative cell lines. P210$^{c\text{-}abl}$ was only detected in the Ph^1-positive cell lines because the *bcr-abl* chimeric gene which encodes P210$^{c\text{-}abl}$ resides on the Ph^1 (4, 5, 11, 13). The level of P210$^{c\text{-}abl}$ was about 4- to 5-fold higher than the level of P145$^{c\text{-}abl}$ in the SK-CML7Bt-33 cell line (Fig. 1*A*, +). The Ph^1-positive erythroid-progenitor cell line K562 (*C*) showed a level of P210$^{c\text{-}abl}$ about 10-fold higher than P145$^{c\text{-}abl}$. However, the level of P210$^{c\text{-}abl}$ was about one-fifth that of P145$^{c\text{-}abl}$ in the Ph^1-positive SK-CML16Bt-1 cell line (Fig. 1*B*, +). Comparison of different autoradiographic exposures roughly indicated that the level of P210$^{c\text{-}abl}$ varies over a 20-fold range between these Ph^1-positive B-cell lines. Analysis of four additional Ph^1-positive B-cell lines demonstrated that the level of P210$^{c\text{-}abl}$ fell into two general classes; some cell lines had a level of P210$^{c\text{-}abl}$ similar to SK-CML7Bt-33 and others had the low level similar to SK-CML16Bt-1 (Table 1). This differs from previous studies with Ph^1-positive myeloid cell lines and patient samples derived from acute-phase CML patients, in which P210$^{c\text{-}abl}$ was detected at a 10-fold higher level than P145$^{c\text{-}abl}$ (refs. 10 and 11; Table 1). There was no large difference in level of chimeric mRNA and P210$^{c\text{-}abl}$ expressed in four myeloid/erythroid-lineage Ph^1-positive cell lines (K562, EM2, EM3, CML22, and BV173; refs. 9 and 11), despite a 4- to 5-fold amplification of *abl*-related sequences in the K562 cell line.

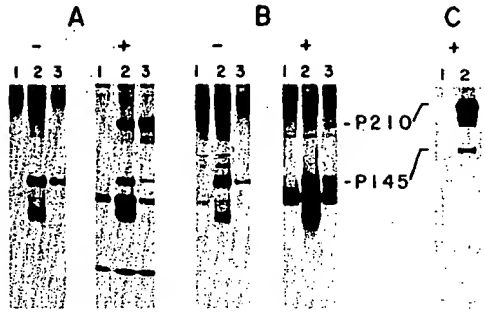

FIG. 1. Detection of variable levels of P210$^{c\text{-}abl}$ in Ph^1-positive B-cell lines. Production of P145$^{c\text{-}abl}$ and P210$^{c\text{-}abl}$ in Epstein–Barr virus-transformed B-cell lines derived from a blast-crisis (*A*) and a chronic-phase (*B*) CML patient was examined by metabolic labeling with [^{32}P]orthophosphate and immunoprecipitation. Ph^1-negative (−) and Ph^1-positive (+) cell lines derived from each patient were analyzed. The Ph^1-negative cell line in *A*,− is SK-CML7BN-2 and in *B*,− is SK-CML16BN-1. The Ph^1-positive cell line in *A*,+ is SK-CML7Bt-33 and in *B*,+ is SK-CML16Bt-1. The K562 cell line, a Ph^1-positive erythroid progenitor cell line spontaneously derived from a blast-crisis patient (33), is represented in *C*. Cells (1.5×10^7) were metabolically labeled with 2 mCi of [^{32}P]orthophosphate for 3–4 hr and then were extracted and clarified by centrifugation. Samples were immunoprecipitated with control normal serum (lanes 1), anti-pEX-2 (lanes 2), or anti-pEX-5 (lanes 3) and analyzed by $NaDodSO_4$/8% PAGE followed by autoradiography with an intensifying screen (3 days for *A* and *C*, 10 days for *B*).

Detection of different levels of P210$^{c\text{-}abl}$ in Fig. 1 could be due to decreased phosphorylation of P210$^{c\text{-}abl}$, a lower level of P210$^{c\text{-}abl}$ synthesis, or altered stability of the protein. To help distinguish among these possibilities, the steady-state level of P210$^{c\text{-}abl}$ in the cell lines was assayed by immunoblotting. The results show that SK-CML7Bt-33 (Fig. 2*A*, +) had a higher level of P210$^{c\text{-}abl}$ than P145, similar to the results with metabolic labeling (Fig. 1). We did not detect P210$^{c\text{-}abl}$ by immunoblotting with 2×10^7 cells of line SK-CML8Bt-3 (Fig. 2*B*, +). Reconstruction experiments using dilutions of cell extracts showed that we could detect about 5–10% the level of P210$^{c\text{-}abl}$ expressed in the K562 cell line (data not shown). We infer that the steady-state level of P210$^{c\text{-}abl}$ in SK-CML8Bt-3 is lower than the level in SK-CML7Bt-33 by a factor of at least 10. The level of P210$^{c\text{-}abl}$ detected in these assays correlated with the amount of P210$^{c\text{-}abl}$ tyrosine kinase activity that could be detected *in vitro* (data not shown).

Different Levels of P210$^{c\text{-}abl}$ Are Reflected in the Amount of Stable *bcr-abl* mRNA. To identify the basis for detection of variable levels of P210$^{c\text{-}abl}$, we examined the production of the *abl* RNA. RNA blot hybridization analysis using a v-*abl* probe (Fig. 3) showed that the normal 6- and 7-kb c-*abl* mRNAs were present at a similar level in Ph^1-positive and -negative cell lines derived from different patients. However, the 8-kb mRNA that encodes P210$^{c\text{-}abl}$ was detected at a 10-fold higher level in SK-CML7Bt-33 (Fig. 3*A*, +) than in SK-CML16Bt-1 (*B*, +), which correlated with the relative level of P210$^{c\text{-}abl}$ detected in each cell line. Analysis of additional cell lines demonstrated that the level of 8-kb RNA directly correlated with the level of P210$^{c\text{-}abl}$ (Table 1). The variation in level of 8-kb RNA detected in these cell lines was not due to loss or gain of Ph^1, because cytogenetic analysis confirmed the presence of Ph^1 in these cell lines (ref. 16 and

Table 1. Relative levels of *bcr-abl* expression in Epstein–Barr virus-immortalized B-cell lines and myeloid CML lines

Cell line*	CML phase†	Ph¹‡	P210§	8-kb mRNA¶
SK-CML7BN-2	BC	−	−	−
SK-CML8BN-10	Chronic	−	−	−
SK-CML8BN-12	Chronic	−	−	−
SK-CML16BN-1	Chronic	−	−	−
SK-CML35BN-1	Chronic	−	−	−
SK-CML7B5-33	BC	+	+++	+++
SK-CML21Bt-1	Acc	+	+++	+++
SK-CML21Bt-6	Acc	+	+++	+++
SK-CML8Bt-3	Chronic	+	+	±
SK-CML16Bt-1	Chronic	+	+	+
SK-CML35Bt-2	Chronic	+	+	+
K562	BC	+	+++++	+++++
BV173	BC	+	+++++	+++++
EM2	BC	+	+++++	+++++

*Cell lines derived from CML patients by transformation with Epstein–Barr virus as described (16). Names of cell lines indicate patient number and Ph¹ status: SK-CML7Bt indicates a cell line derived from patient 7 that carries the 9;22 Ph¹ translocation; N indicates a normal karyotype. Myeloid-erythroid cell lines (K562, EM2, and BV173) are described in previous publications (9, 11, 22, 33).

†Status of patient at the time cell line was derived. BC, blast crisis; Acc, accelerated phase.

‡Presence (+) or absence (−) of Ph¹ as demonstrated by karyotypic or Southern blot analysis.

§P210$^{c\text{-}abl}$ detected as described in legend to Fig. 1. B-cell lines derived from blast-crisis and accelerated-phase patients had levels of P210 3- to 5-fold higher (+++) than levels of P145. Chronic-phase-derived cell lines had P210 levels lower than or just equivalent (+) to the level of P145. Myeloid and erythroid lines had levels of P210 5- to 10-fold higher than P145 (+++++).

¶Eight-kilobase *bcr-abl* mRNA detected as described in legend to Fig. 2. Symbols: ±, borderline detectable; +++++, level of 8-kb mRNA 5- to 10-fold higher than that of the 6- and 7-kb c-*abl* mRNA species; +++, level of 8-kb mRNA 3- to 5-fold higher than that of the 6- and 7-kb species; +, a level approximately equivalent to that of the 6- and 7-kb messages.

data not shown). There was no difference in the copy number of *abl*-related sequences as judged by Southern blot analysis (Fig. 4). Only the K562 cell line control showed an amplification of *abl* sequences, as previously reported (22, 23). These combined data suggest that differential *bcr-abl* mRNA expression from a single gene template is responsible for the variable levels of P210$^{c\text{-}abl}$ detected. This could be mediated

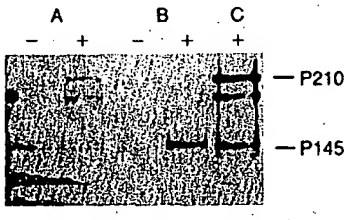

FIG. 2. Analysis of steady-state *abl* protein levels by immunoblotting. Cell extracts prepared from 2×10^7 cells of lines SK-CML7BN-2 (*A*,−), SK-CML7Bt-33 (*A*,+), SK-CML8BN-10 (*B*,−), and SK-CML8Bt-3 (*B*,+) were concentrated by immunoprecipitation with anti-pEX-2 plus anti-pEX-5. Samples were then electrophoresed in a NaDodSO$_4$/8% polyacrylamide gel and transferred to nitrocellulose, using protease-facilitated transfer (18). *abl* proteins were detected using a mixture of two monoclonal antibodies directed against the pEX-2 and pEX-5 *abl*-protein fragments produced in bacteria (19) as a probe and a peroxidase-conjugated goat anti-mouse second-stage antibody (Bio-Rad) for development.

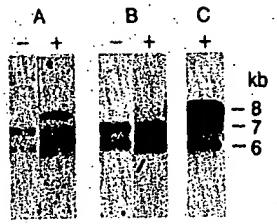

FIG. 3. Comparison of *abl* RNA levels in Ph¹-positive and -negative B-cell lines. The levels of the normal 6- and 7-kb c-*abl* RNAs and the 8-kb *bcr-abl* RNA were analyzed by blot hybridization using a v-*abl* probe. RNA was extracted from Ph¹-negative lines SK-CML7BN-2 (*A*,−) and SK-CML16BN-1 (*B*,−), from Ph¹-positive lines SK-CML6Bt-33 (*A*,+) and SK-CML16Bt-3 (*B*,+), and from line K562 (*C*,+) by the NaDodSO$_4$/urea/phenol method (20). Polyadenylylated RNA was purified by oligo(dT) affinity chromatography, and 15 μg of each sample was electrophoresed in a 1% agarose/formaldehyde gel and then transferred to nitrocellulose. The blotted RNAs were hybridized with a nick-translated v-*abl* fragment probe (21) and then autoradiographed for 4 days.

by factors influencing the transcription rate of the *bcr-abl* gene or the stability of the mRNA.

DISCUSSION

Several lines of evidence suggest that formation of Ph¹ is not the primary event that affects the stem cell in CML. Patients have been identified that present with the clinical picture of CML but only later develop Ph¹ (1). This observation, coupled with studies of *G6PD* (glucose-6-phosphate dehydrogenase)-heterozygous females with CML that demonstrate stem-cell clonality by isozyme analysis among cell

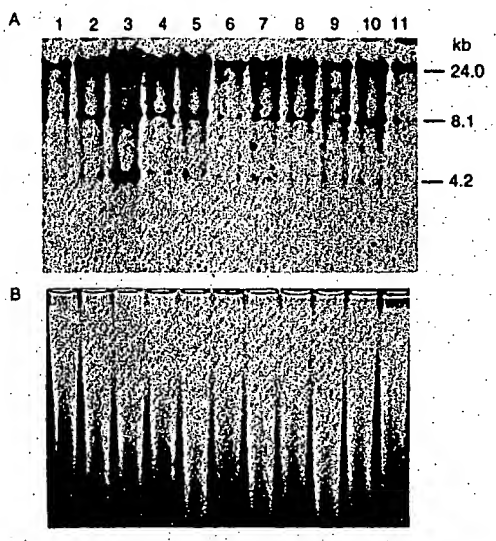

FIG. 4. Southern blot analysis of *abl* sequences in Ph¹-positive and -negative B-cell lines. High molecular weight DNA (15 μg) was digested with restriction endonuclease *Bam*HI, separated in a 0.8% agarose gel, and then transferred to nitrocellulose. The blotted DNA fragments were hybridized with a nick-translated, 2.4-kb *Bgl* II v-*abl* fragment (1.5×10^8 cpm/μg; ref. 21) and exposed for 4 days. (*A*) Autoradiogram of *abl*-specific fragments in cell lines HL-60 (lane 1), EM2 (lane 2), K562 (lane 3), SK-CML7Bt-33 (lane 4), SK-CML8Bt-3 (lane 5), SK-CML16Bt-1 (lane 6), SK-CML21Bt-6 (lane 7), SK-CML35Bt-2 (lane 8), SK-CML7BN-2 (lane 9), SK-CML8BN-2 (lane 10), and SK-CML35BN-1 (lane 11). (*B*) Ethidium bromide staining of agarose gel prior to transfer to nitrocellulose, showing the level of variation in amount of DNA loaded per lane.

populations that lack the Ph^1 marker, supports a secondary or complementary role for Ph^1 in the progression of the disease (24, 25). This chromosome marker is found in chronic, accelerated, and blast-crisis phases of the disease. It is likely that Ph^1 confers some growth advantage, since cells with the marker chromosome eventually predominate the marrow and peripheral blood even in chronic phase. During the phase of blast crisis, many patients develop additional chromosome abnormalities, including duplication of Ph^1, a variety of trisomies, and complex translocations (26). This is suggestive evidence for Ph^1 being a necessary but not sufficient genetic change for the full evolution of the disease.

The realization that one molecular result of Ph^1 is the generation of a chimeric *bcr-abl* protein with functional characteristics and structure analogous to the *gag-abl* transforming protein of the Abelson murine leukemia virus strengthens the argument for an important role of Ph^1 in the pathogenesis of CML. Although the Abelson virus is generally considered a rapidly transforming retrovirus, its effects can range from overcoming growth factor requirements, to cellular lethality, to induction of highly oncogenic tumors in a number of hematopoietic cell lineages (27, 28). Even in the transformation of murine cell targets, there are several lines of evidence that suggest that the growth-promoting activity of the v-*abl* gene product is complemented by further cellular changes in the production of the malignant-cell phenotype (29–31).

The regulation of *bcr-abl* gene expression is complex because the 5′ end of the gene is derived from the non-*abl* sequences, *bcr*, normally found on chromosome 22 (6). The level of stable message for the normal *bcr* gene and the normal *abl* gene are both much lower than the level of the *bcr-abl* message and protein from cell lines and clinical specimens derived from myeloid blast-crisis patients (5, 6, 11). Therefore, the high level of *bcr-abl* expression cannot simply be attributed to the regulatory sequences associated with *bcr*. Possibly, creation of the chimeric gene disrupts the normal regulatory sequences and results in a higher level of expression. Variation in *bcr-abl* expression may result from secondary changes in the structure of the chimeric gene or function of *trans*-acting factors that occur during evolution of the disease. Our analysis of $P210^{c\text{-}abl}$ and the 8-kb mRNA in Epstein–Barr virus-transformed Ph^1-positive B-cell lines demonstrates that stable message and protein levels from the *bcr-abl* gene can vary over a wide range. This variation does not result from a change in the number of *bcr-abl* templates secondary to gene amplification but more likely from changes in either transcription rate or mRNA stability. We suspect this range of *bcr-abl* expression is not limited to lymphoid cells. Analysis of peripheral blood leukocytes derived from an unusual CML patient who has been in chronic phase with myeloid predominance for 16 years showed a level of $P210^{c\text{-}abl}$ one-fifth that of $P145^{c\text{-}abl}$, as detected by metabolic labeling with [^{32}P]orthophosphate and immunoprecipitation (S.C., O.N.W., and P. Greenberg, unpublished observations). Lower levels of expression of the chimeric mRNA have been demonstrated in clinical samples from chronic-phase CML patients compared to acute-phase CML patients (9). Others have reported chronic-phase patients with variable but, in some cases, relatively high levels of the *bcr-abl* mRNA (32). The sampling variation and the heterogenous mixture of cell types in clinical samples complicate such analyses. Further work is needed to evaluate whether there is a defined change in $P210^{c\text{-}abl}$ expression during the progression of CML. It is interesting to note that among the limited sample of Ph^1-positive B-cell lines we have examined (Table 1), we have seen higher levels of $P210^{c\text{-}abl}$ in those derived from patients at more advanced stages of the disease. It will be important to search for cell-type-specific mechanisms that might regulate expression of *bcr-abl* from Ph^1.

We thank Bonnie Hechinger and Carol Crookshank for excellent secretarial assistance and Margaret Newman for excellent technical assistance. This work was supported by grants from the National Institutes of Health (to O.N.W. and B.C.). J.B.K. was supported as a predoctoral fellow on the Public Health Service Cellular and Molecular Biology Training Grant GM07185. S.C. is a postdoctoral fellow of the Leukemia Society of America.

1. Champlin, R. E. & Golde, D. W. (1985) *Blood* **65**, 1039–1047.
2. Rowley, J. D. (1973) *Nature (London)* **243**, 290–291.
3. Heisterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartram, C. R. & Grosveld, G. (1983) *Nature (London)* **306**, 239–242.
4. Bartram, C. R., de Klein, A., Hagemeijer, A., van Agthoven, T., van Kessel, A. G., Bootsma, D., Grosveld, G., Ferguson-Smith, M. A., Davies, T., Stone, M., Heisterkamp, N., Stephenson, J. R. & Groffen, J. (1983) *Nature (London)* **306**, 277–280.
5. Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, D. (1985) *Nature (London)* **315**, 550–554.
6. Heisterkamp, N., Stam, K. & Groffen, J. (1985) *Nature (London)* **315**, 758–761.
7. Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R. & Grosveld, G. (1984) *Cell* **36**, 93–99.
8. Gale, R. P. & Canaani, E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5648–5652.
9. Collins, S., Kubonishi, L., Miyoshi, I. & Groudine, M. T. (1984) *Science* **225**, 72–74.
10. Konopka, J. B., Watanabe, S. M. & Witte, O. N. (1984) *Cell* **7**, 1035–1042.
11. Konopka, J. B., Watanabe, S. M., Singer, J., Collins, S. & Witte, O. N. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1810–1814.
12. Kloetzer, W., Kurzrock, R., Smith, L., Talpaz, M., Spiller, M., Gutterman, J. & Arlinghaus, R. (1985) *Virology* **140**, 230–238.
13. Kozbor, D., Giallongo, A., Sierzega, M. E., Konopka, J. B., Witte, O. N., Showe, L. C. & Croce, C. M. (1985) *Nature (London)*, in press.
14. Davis, R. L., Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* **5**, 204–213.
15. Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell. Biol.* **5**, 3116–3123.
16. Nitta, M., Kato, Y., Strife, A., Wachter, M., Fried, J., Perez, A., Jhanwar, S., Duigou, R., Chaganti, R. S. K. & Clarkson, B. (1985) *Blood* **66**, 1053–1061.
17. Konopka, J. B., Davis, J. L., Watanabe, S. M., Ponticelli, A. S., Schiff-Maker, L., Rosenberg, N. & Witte, O. N. (1984) *Virology* **51**, 223–232.
18. Gibson, W. (1981) *Anal. Biochem.* **118**, 1–3.
19. Schiff-Maker, L., Konopka, J. B., Clark, S., Witte, O. N. & Rosenberg, N. (1986) *J. Virol.* **57**, 1182–1186.
20. Schwartz, R. C., Sonenshein, G. E., Bothwell, A. & Gefter, M. L. (1981) *J. Immunol.* **126**, 2104–2108.
21. Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. (1980) *Cell* **22**, 777–785.
22. Collins, S. J. & Groudine, M. T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4813–4817.
23. Selden, J. R., Emanuel, B. S., Wang, E., Cannizzaro, L., Palumbo, A., Erikson, J., Nowell, P. C., Rovera, G. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7289–7292.
24. Fialkow, P. J., Martin, P. J., Najfeld, V., Penfold, G. K., Jacobson, R. J. & Hansen, J. A. (1981) *Blood* **58**, 158–163.
25. Martin, P. J., Najfeld, V. & Fialkow, P. J. (1982) *Can. Gen. Cytogenet.* **6**, 359–368.
26. Rowley, J. D. (1980) *Annu. Rev. Genet.* **14**, 17–40.
27. Whitlock, C. A. & Witte, O. N. (1984) *Adv. Immunol.* **37**, 74–98.
28. Pierce, J. H., Di Fiore, P. P., Aaronson, S. A., Potter, M., Pumphrey, J., Scott, A. & Ihle, J. N. (1985) *Cell* **41**, 685–693.
29. Whitlock, C. A., Ziegler, S. & Witte, O. N. (1983) *Mol. Cell. Biol.* **3**, 596–604.
30. Wolf, D., Harris, N. & Rotter, V. (1984) *Cell* **38**, 119–126.
31. Klein, G. & Klein, E. (1985) *Nature (London)* **315**, 190–195.
32. Stam, K., Jr., Heisterkamp, N., Grosveld, G., de Klein, A., Verma, R., Colemen, M., Dosik, H. & Groffen, J. (1985) *N. Engl. J. Med.* **313**, 1429–1433.
33. Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* **45**, 321–334.

Aneuploidy and cancer

Subrata Sen, PhD

Numeric aberrations in chromosomes, referred to as aneuploidy, is commonly observed in human cancer. Whether aneuploidy is a cause or consequence of cancer has long been debated. Three lines of evidence now make a compelling case for aneuploidy being a discrete chromosome mutation event that contributes to malignant transformation and progression process. First, precise assay of chromosome aneuploidy in several primary tumors with *in situ* hybridization and comparative genomic hybridization techniques have revealed that specific chromosome aneusomies correlate with distinct tumor phenotypes. Second, aneuploid tumor cell lines and *in vitro* transformed rodent cells have been reported to display an elevated rate of chromosome instability, thereby indicating that aneuploidy is a dynamic chromosome mutation event associated with transformation of cells. Third, and most important, a number of mitotic genes regulating chromosome segregation have been found mutated in human cancer cells, implicating such mutations in induction of aneuploidy in tumors. Some of these gene mutations, possibly allowing unequal segregations of chromosomes, also cause tumorigenic transformation of cells *in vitro*. In this review, the recent publications investigating aneuploidy in human cancers, rate of chromosome instability in aneuploidy tumor cells, and genes implicated in regulating chromosome segregation found mutated in cancer cells are discussed. Curr Opin Oncol 2000, 12:82–88

The University of Texas, M.D. Anderson Cancer Center, Department of Laboratory Medicine, Houston, Texas, USA

Correspondence to Subrata Sen, PhD, The University of Texas, MD Anderson Cancer Center, Department of Laboratory Medicine, Box 054, 1515 Holcombe Blvd., Houston, TX 77030, USA; tel: 713-792-2560; fax: 713-792-4094; e-mail: ssen@mdanderson.org

Current Opinion in Oncology 2000 12:82–88

Abbreviations

CGH	comparative genomic hybridization
CHE	Chinese hamster embryo cells
FISH	fluorescence in situ hybridization
HPRC	hereditary papillary renal carcinoma
ISH	*in situ* hybridization

ISSN 1040-8746

Cancer research over the past decade has firmly established that malignant cells accumulate a large number of genetic mutations that affect differentiation, proliferation, and cell death processes. In addition, it is also recognized that most cancers are clonal, although they display extensive heterogeneity with respect to karyotypes and phenotypes of individual clonal populations. It is estimated that numeric chromosomal imbalance, referred to as *aneuploidy*, is the most prevalent genetic change recorded among over 20,000 solid tumors analyzed thus far [1]. Phenotypic diversity of the clonal populations in individual tumors involve differences in morphology, proliferative properties, antigen expression, drug sensitivity, and metastatic potentials. It has been proposed that an underlying acquired genetic instability is responsible for the multiple mutations detected in cancer cells that lead to tumor heterogeneity and progression [2]. In a somewhat contradictory argument, it has also been suggested that clonal expansion due to selection of cells undergoing normal rates of mutation can explain malignant transformation and progression process in humans [3]. Acquired genetic instability, nonetheless, is considered important for more rapid progression of the disease [4••]. Although the original hypothesis on genetic instability in cancer primarily focused on chromosome imbalances in the form of aneuploidy in tumor cells, the actual relevance of such mutations in cancer remains a controversial issue.

Whether or not aneuploidy contributes to the malignant transformation and progression process has long been debated. A prevalent idea on genetics of cancer referred to as "somatic gene mutation hypothesis" contends that gene mutations at the nucleotide level alone can cause cancer by either activating cellular proto-oncogenes to dominant cancer causing oncogenes and/or by inactivating growth inhibitory tumor suppressor genes. In this scheme of things chromosomal instability in the form of aneuploidy is a mere consequence rather than a cause of malignant transformation and progression process.

In this review, some of the recent observations on the subject are discussed and compelling evidence is provided to suggest that aneuploidy is a distinct form of genetic instability in cancer that frequently correlates with specific phenotypes and stages of the disease. Furthermore, discrete genetic targets affecting chromosomal stability in cancer cells, recently identified, are also discussed. These data provide a new direction toward elucidating the molecular mechanisms responsi-

ble for induction of aneuploidy in cancer and may eventually be exploited as novel therapeutic targets in the future.

Genetic alterations in cancer

Alterations in many genetic loci regulating growth, senescence, and apoptosis, identified in tumor cells, have led to the current understanding of cancer as a genetic disease. The genetic changes identified in tumors include: subtle mutations in genes at the nucleotide level; chromosomal translocations leading to structural rearrangements in genes; and numeric changes in either partial segments of chromosomes or whole chromosomes (aneuploidy) causing imbalance in gene dosage.

For the purpose of this review, both segmental and whole chromosome imbalances leading to altered DNA dosage in cancer cells are included as examples of aneuploidy.

Incidence of aneuploidy in cancer

Evidence of aneuploidy involving one or more chromosomes have been commonly reported in human tumors. Although these observations were initially made using classic cytogenetic techniques late in a tumor's evolution and were difficult to correlate with cancer progression, more recent studies have reported association of specific nonrandom chromosome aneuploidy with different biologic properties such as loss of hormone dependence and metastatic potential [5].

Classic cytogenetic studies performed on tumor cells had serious limitations in scope because they were applicable only to those cases in which mitotic chromosomes could be obtained. Because of low spontaneous rates of cell division in primary tumors, analyses depended on cells either derived selectively from advanced metastases or those grown *in vitro* for variable periods of time. In both instances, metaphases analyzed represented only a subset of primary tumor cell population. Two major advances in cytogenetic analytic techniques, *in situ* hybridization (ISH) and comparative genomic hybridization (CGH), have allowed better resolution of chromosomal aberrations in freshly isolated tumor cells [6]. ISH analyses with chromosome-specific DNA probes, a powerful adjunct to metaphasic analysis, allows assessment of chromosomal anomalies within tumor cell populations in the contexts of whole nuclear architecture and tissue organization. CGH allows genome wide screening of chromosomal anomalies without the use of specific probes even in the absence of prior knowledge of chromosomes involved. Although both techniques have certain limitations in terms of their resolution power, they nonetheless provide a better approximation of chromosomal changes occurring among tumors of various histology, grade, and stage compared with what was possible with the classic cytogenetic techniques. Genomic ploidy measurements have also been performed at the DNA level with flow cytometry and cytofluorometric methods. Although these assays underestimate chromosome ploidy due to a chromosomal gain occasionally masking a chromosomal loss in the same cell, several studies using these methods have supported the conclusion that DNA aneuploidy closely associates with poor prognosis in various cancers [7,8]. This discussion of some recent examples published on aneuploidy in cancer includes discussion of studies dealing with DNA ploidy measurements as well. Most of these observations are correlative without direct proof of specific involvement of genes on the respective chromosomes. Identification of putative oncogenes and tumor suppressor genes on gained and lost chromosomes in aneuploid tumors, however, are providing strong evidence that chromosomes involved in aneuploidy play a critical role in the tumorigenic process.

In renal tumors, either segmental or whole chromosome aneuploidy appears to be uniquely associated with specific histologic subtypes [9]. Tumors from patients with hereditary papillary renal carcinomas (HPRC) commonly show trisomy of chromosome 7, when analyzed by CGH. Germline mutations of a putative oncogene *MET* have been detected in patients with HPRC. A recent study [10] has demonstrated that an extra copy of chromosome 7 results in nonrandom duplication of the mutant *MET* allele in HPRC, thereby implicating this trisomy in tumorigenesis. The study suggested that mutation of *MET* may render the cells more susceptible to errors in chromosome replication, and that clonal expansion of cells harboring duplicated chromosome 7 reflects their proliferative advantage. In addition to chromosome 7, trisomy of chromosome 17 in papillary tumors and also of chromosome 8 in mesoblastic nephroma are commonly seen. Association of specific chromosome imbalances with benign and malignant forms of papillary renal tumors, therefore, not only contribute to an understanding of tumor origins and evolution, but also implicate aneuploidy of the respective chromosomes in the tumorigenic transformation process.

In colorectal tumors, chromosome aneuploidy is a common occurrence. In fact, molecular allelotyping studies have suggested that limited karyotyping data available from these tumors actually underestimate the true extent of these changes. Losses of heterozygosity reflecting loss of the maternal or paternal allele in tumors are widespread and often accompanied by a gain of the opposite allele. Therefore, for example, a tumor could lose a maternal chromosome while duplicating the same paternal chromosome, leaving the tumor cell

with a normal karyotype and ploidy but an aberrant allelotype. It has been estimated that cancer of the colon, breast, pancreas, or prostate may lose an average of 25% of its alleles. It is not unusual to discover that a tumor has lost over half of its alleles [4]. In clinical settings, DNA ploidy measurements have revealed that DNA aneuploidy indicates high risk of developing severe premalignant changes in patients with ulcerative colitis, who are known to have an increased risk of developing colorectal cancer [11]. DNA aneuploidy has been found to be one of the useful indicators of lymph node metastasis in patients with gastric carcinoma and associated with poor outcome compared with diploid cases [12,13]. CGH analyses of chromosome aneuploidy, on the other hand, was reported to correlate gain of chromosome 20q with high tumor S phase fractions and loss of 4q with low tumor apoptotic indices [14]. Aneuploidy of chromosome 4 in metastatic colorectal cancer has recently been confirmed in studies that used unbiased DNA fingerprinting with arbitrarily primed polymerase chain reactions to detect moderate gains and losses of specific chromosomal DNA sequences [15]. The molecular karyotype (amplotype) generated from colorectal cancer revealed that moderate gains of sequences from chromosomes 8 and 13 occurred in most tumors, suggesting that overrepresentation of these chromosomal regions is a critical step for metastatic colorectal cancer.

In addition to being implicated in tumorigenesis and correlated with distinct tumor phenotypes, chromosome aneuploidy has been used as a marker of risk assessment and prognosis in several other cancers. The potential value of aneuploidy as a noninvasive tool to identify individuals at high risk of developing head and neck cancer appears especially promising. Interphase fluorescence *in situ* hybridization (FISH) revealed extensive aneuploidy in tumors from patients with head and neck squamous cell carcinomas (HNSCC) and also in clinically normal distant oral regions from the same individuals [16,17]. It has been proposed that a panel of chromosome probes in FISH analyses may serve as an important tool to detect subclinical tumorigenesis and for diagnosis of residual disease. The presence of aneuploid or tetraploid populations is seen in 90% to 95% of esophageal adenocarcinomas, and when seen in conjunction with Barrett's esophagus, a premalignant condition, predicts progression of disease [18,19]. Chromosome ploidy analyses in conjunction with loss of heterozygosity and gene mutation studies in Barrett's esophagus reflect evolution of neoplastic cell lineages *in vivo* [20]. Evolution of neoplastic progeny from Barrett's esophagus following somatic genetic mutations frequently involves bifurcations and loss of heterozygosity at several chromosomal loci leading to aneuploidy and cancer. Accordingly, it is hypothesized that during tumor cell evolution diploid cell progenitors with somatic genetic abnormalities undergo expansion with acquired genetic instability. Such instability, often manifested in the form of increased incidence of aneuploidy, enters a phase of clonal evolution beginning in premalignant cells that proceeds over a period of time and occasionally leads to malignant transformation. The clonal evolution continues even after the emergence of cancer.

The significance of DNA and chromosome aneuploidy in other human cancers continue to be evaluated. Among papillary thyroid carcinomas, aneuploid DNA content in tumor cells was reported to correlate with distant metastases, reflecting worsened prognosis [21]. Genome wide screening of follicular thyroid tumors by CGH, on the other hand, revealed frequent loss of chromosome 22 in widely invasive follicular carcinomas [22]. Chromosome copy number gains in invasive neoplasm compared with foci of ductal carcinoma *in situ* (DCIS) with similar histology have been proposed to indicate involvement of aneuploidy in progression of human breast cancer [23]. ISH analyses of cervical intraepithelial neoplasia has provided suggestive evidence that chromosomes 1, 7 and X aneusomy is associated with progression toward cervical carcinoma [24].

Although the prognostic value of numeric aberrations remains a matter of debate in human hematopoietic neoplasia, there have been recent studies to suggest that the presence of monosomy 7 defines a distinct subgroup of acute myeloid leukemia patients [25]. It is interesting in this context that therapy-related myelodysplastic syndromes have been reported to display monosomy 5 and 7 karyotypes, reflecting poor prognosis [26].

The clinical observations, mentioned previously, are supported by *in vitro* studies in human and rodent cells in which aneuploidy is induced at early stages of transformation [27,28]. It is even suggested that aneuploidy may cause cell immortalization, in some instances, that is a critical step preceeding transformation.

Finally, in an interesting study to develop transgenic mouse models of human chromosomal diseases, chromosome segment specific duplication and deletions of the genome were reported to be constructed in mouse embryonic stem cells [29]. Three duplications for a portion of mouse chromosome 11 syntenic with human chromosome 17 were established in the mouse germline. Mice with 1Mb duplication developed corneal hyperplasia and thymic tumors. The findings represent the first transgenic mouse model of aneuploidy of a defined chromosome segment that documents the direct role of chromosome aneusomy in tumorigenesis.

Aneuploidy as "dynamic cancer-causing mutation" instead of a "consequential state" in cancer

According to the hypothesis previously discussed, aneuploidy represents either a "gain of function" or "loss of function" mutation at the chromosome level with a causative influence on the tumorigenesis process. The hypothesis, however, is based only on circumstantial evidence even though existence of aneuploidy is correlated with different tumor phenotypes. The existence of numeric chromosomal alterations in a tumor does not mean that the change arose as a dynamic mutation due to genomic instability, because several factors could lead to consequential aneuploidy in tumors, also. Although aneuploidy as a dynamic mutation due to genomic instability in tumor cells would occur at a certain measurable rate per cell generation, a consequential state of aneuploidy in tumors may not occur at a predictable rate under similar conditions or in tumors with similar phenotypes. In addition to genomic instability, differences in environmental factors with selective pressure, could explain high incidence of aneuploidy and other somatic mutations in tumors compared with normal cells [4]. These include humoral, cell substratum, and cell-cell interaction differences between tumor and normal cell environments. It could be argued that despite similar rates of spontaneous aneuploidy induction in normal and tumor cells, the latter are selected to proliferate due to altered selective pressure in the tumor cell environment, whereas the normal cells are eliminated through activation of apoptosis. Alternatively, of course, one could postulate that selective expression or overexpression of anti-apoptotic proteins or inactivation of proapoptotic proteins in tumor cells may counteract default induction of apoptosis in G2/M phase cells undergoing missegregation of chromosomes. Recent demonstration of overexpression of a G2/M phase anti-apoptotic protein survivin in cancer cells [30] suggests that this protein may favor aberrant progression of aneuploid transformed cells through mitosis. This would then lead to proliferation of aneuploid cell lineages, which may undergo clonal evolution.

To ascertain that aneuploidy is a dynamic mutational event, various human tumor cell lines and transformed rodent cell lines have been analyzed for the rate of aneuploidy induction. When grown under controlled *in vitro* conditions, such conditions ensure that environmental factors do not influence selective proliferation of cells with chromosome instability. In one study, Lengauer *et al.* [31•] provided unequivocal evidence by FISH analyses that losses or gains of multiple chromosomes occurred in excess of 10^{-2} per chromosome per generation in aneuploid colorectal cancer cell lines. The study further concluded that such chromosomal instability appeared to be a dominant trait. Using another *in vitro* model system of Chinese hamster embryo (CHE) cells, Duesberg *et al.* [32•] have also obtained similar results. With clonal cultures of CHE cells, transformed with nongenotoxic chemicals and a mitotic inhibitor, these authors demonstrated that the overwhelming majority of the transformed colonies contained more than 50% aneuploid cells, indicating that aneuploidy would have originated from the same cells that underwent transformation. All the transformed colonies tested were tumorigenic. It was further documented that the ploidy factor representing the quotient of the modal chromosome number divided by the normal diploid number, in each clone, correlated directly with the degree of chromosomal instability. Therefore, chromosomal instability was found proportional to the degree of aneuploidy in the transformed cells and the authors hypothesized that aneuploidy is a unique mechanism of simultaneously altering and destabilizing, in a massive manner, the normal cellular phenotypes. In the absence of any evidence that the transforming chemicals used in the study did not induce other somatic mutations, it is difficult to rule out the contribution of such mutations in the transformation process. These results nonetheless make a strong case for aneuploidy being a dynamic chromosome mutation event intimately associated with cancer.

Aneuploidy versus somatic gene mutation in cancer

The idea that numeric chromosome imbalance or aneuploidy is a direct cause of cancer was proposed at the turn of the century by Theodore Boveri [33]. However, the hypothesis was largely ignored over the last several decades in favor of the somatic gene mutation hypothesis, mentioned earlier. Evidence accumulating in the literature lately on specific chromosome aneusomies recognized in primary tumors, incidence of aneuploidy in cells undergoing transformation, and aneuploid tumor cells showing a high rate of chromosome instability have led to the rejuvenation of Boveri's hypothesis. The concept has recently been discussed as a "vintage wine in a new bottle" [34•]. The author points out that except for rare cancers caused by dominant retroviral oncogenes, diploidy does not seem to occur in solid tumors, whereas aneuploidy is a rule rather than exception in cancer.

Aneuploidy as an effective mutagenic mechanism driving tumor progression, on the other hand, is being recognized as a viable solution to the paradox that with known mutation rate in non-germline cells ($\sim 10^{-7}$ per gene per cell generation) tumor cell lineages cannot accumulate enough mutant genes during a human lifetime [35]. The concept is gaining significant credibility since genes that potentially affect chromosome segregation were found mutated in human cancer. Some of

these genes have also been shown to have transforming capability in *in vitro* assays. Selected recent publications describing the findings are being discussed below in reference to the mitotic targets potentially involved in inducing chromosome segregation anomalies in cells.

Potential mitotic targets and molecular mechanisms of aneuploidy

Because aneuploidy represents numeric imbalance in chromosomes, it is reasonable to expect that aneuploidy arises due to missegregation of chromosomes during cell division. There are many potential mitotic targets, which could cause unequal segregation of chromosomes (Fig. 1). Recent investigations have identified several genes involved in regulating these mitotic targets and mitotic checkpoint functions, which can be implicated in induction of aneuploidy in tumor cells. This discussion is restricted to those mitotic targets and checkpoint genes whose abnormal functioning has been observed in cancer or has been shown to cause tumorigenic transformation of cells, in recent years. The role of telomeres is discussed elsewhere in this issue. For a more detailed description of the components of mitotic machinery and their possible involvement in causing chromosome segregation abnormalities in tumor cells, readers may refer to a recently published review [36•].

Among the mitotic targets implicated in cancer, centrosome defects have been observed in a wide variety of malignant human tumors. Centrosomes play a central role in organizing the microtubule network in interphase cells and mitotic spindle during cell division. Multipolar mitotic spindles have been observed in human cancers *in situ* and abnormalities in the form of supernumerary centrosomes, centrosomes of aberrant size and shape as well as aberrant phosphorylation of centrosome proteins have been reported in prostate, colon, brain, and breast tumors [37,38]. In view of the findings that abnormal centrosomes retain the ability to nucleate microtubules *in vitro*, it is conceivable that cells with abnormal centrosomes may missegregate chromosomes producing aneuploid cells. The molecular and genetic bases of abnormal centrosome generation and the precise pathway through which they regulate the chromosome segregation process remain to be elucidated. Recent discovery of a centrosome-associated kinase STK15/BTAK/aurora2, naturally amplified and overexpressed in human cancers, has raised the interesting possibility that aberrant expression of this kinase is critically involved in abnormal centrosome function and unequal chromosome segregation in tumor cells [39,40]. Exogenous expression of the kinase in rodent and human cells was found to correlate with an abnormal number of centrosomes, unequal partitioning of chromosomes during division, and tumorigenic transformation of cells. It is relevant in this context to mention that the Xenopus homologue of human STK15/BTAK/aurora2 kinase has recently been shown to phosphorylate a microtubule motor protein XlEg5, the human orthologue of which is known to participate in the centrosome separation during mitosis [41]. Findings on STK15/aurora2 kinase, thus, provide an interesting lead to a possible molecular mechanism of centrosome's role in oncogenesis. Centrosomes have, of late, been implicated in oncogenesis from studies revealing supernumerary centrosomes in *p53*-deficient fibroblasts and overexpression of another centrosome kinase PLK1 being detected in human non-small cell lung cancer [42].

Figure 1. Potential mitotic targets causing aneuploidy in oncogenesis

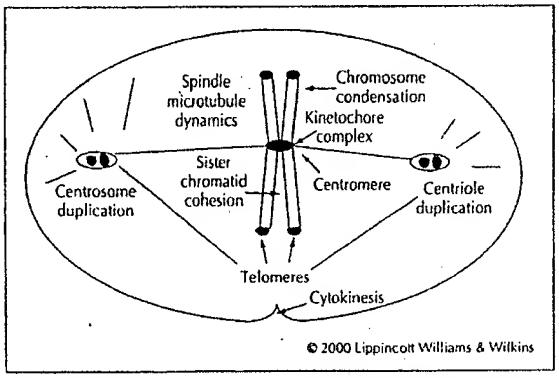

Diagram illustrates that defects in several processes involving chromosomal, spindle microtubule, and centrosomal targets, in addition to abnormal cytokinesis, may cause unequal partitioning of chromosomes during mitosis, leading to aneuploidy. Recently obtained evidence in favor of some of these possibilities is discussed in the text.

One of the critical events that ensures equal partitioning of the chromosomes during mitosis is the proper and timely separation of sister chromatids that are attached to each other and to the mitotic spindle. Untimely separation of sister chromatids has been suspected as a cause of aneuploidy in human tumors. Cohesion between sister chromatids is established during replication of chromosomes and is retained until the next metaphase/anaphase transition. It has been shown that during metaphase-anaphase transition, the anaphase promoting complex/cyclosome triggers the degradation of a group of proteins called securins that inhibit sister chromatid separation. A vertebrate securin (v-securin) has recently been identified that inhibits sister chromatid separation and is involved in transformation and tumorigenesis. Subsequent analysis revealed that the human securin is identical to the product of the gene called pituitary tumor transforming gene, which is overexpressed in some tumors and exhibits transforming activity in NIH3T3 cells. It is proposed that elevated expression of the v-securin may contribute to generation of malignant tumors due to

chromosome gain or loss produced by errors in chromatid separation [43•].

Normal progression through mitosis during prophase to anaphase transition is monitored at least at two checkpoints. One checkpoint operates during early prophase at G2 to metaphase progression while the second ensures proper segregation of chromosomes during metaphase to anaphase transition. Several mitotic checkpoint genes responding to mitotic spindle defects have been identified in yeast. The metaphase-anaphase transition is delayed following activation of this checkpoint during which kinetochores remain unattached to the spindle. The signal is transmitted through a kinetochore protein complex consisting of Mps1p and several Mad and Bub proteins [44]. It is expected that for unequal chromosome segregation to be perpetuated through cell proliferation cycles giving rise to aneuploidy, checkpoint controls have to be abrogated.

Following this logic, Vogelstein *et al.* [45•] hypothesized that aneuploid tumors would reveal mutation in mitotic spindle checkpoint genes. Subsequent studies by these investigators have proven the validity of this hypothesis and a small fraction of human colorectal cancers have revealed the presence of mutations in either hBub1 or hBubR1 checkpoint genes. It was further revealed that mutant BUB1 could function in a dominant negative manner conferring an abnormal spindle checkpoint when expressed exogenously. Inactivation of spindle checkpoint function in virally induced leukemia has also recently been documented following the finding that hMAD1 checkpoint protein is targeted by the Tax protein of the human T-cell leukemia virus type 1. Abrogation of hMAD1 function leads to multinucleation and aneuploidy [46].

In addition to mitotic spindle checkpoint defects, failed DNA damage checkpoint function in yeast is frequently associated with aberrant chromosome segregation as well. It, therefore, appears intriguing yet relevant that the human *BRCA1* gene, proposed to be involved in DNA damage checkpoint function, when mutated by a targeted deletion of exon 11 led to defective G2/M cell cycle checkpoint function and genetic instability in mouse embryonic fibroblasts [47]. The cells revealed multiple functional centrosomes and unequal chromosome segregation and aneuploidy. Although the molecular basis for these abnormalities is not known at this time, it raises the interesting possiblility that such an aneuploidy-driven mechanism may be involved in tumorigenesis in individuals carrying germline mutations of *BRCA1* gene.

Conclusion

Growing evidence from human tumor cytogenetic investigations strongly suggest that aneuploidy is associated with the development of tumor phenotypes. Clinical findings of correlation between aneuploidy and tumorigenesis are supported by studies with *in vitro* grown transformed cell lines. Molecular genetic analyses of tumor cells provide credible evidence that mutations in genes controlling chromosome segregation during mitosis play a critical role in causing chromosome instability leading to aneuploidy in cancer. Further elucidation of molecular and physiologic bases of chromosome instability and aneuploidy induction could lead to the development of new therapeutic approaches for common forms of cancer.

Acknowledgments

The author is thankful to Drs. Bill Brinkley and Pramila Sen for discussions and advice. Help from Ms. Donna Sprabary and Ms. Hongyi Zhou in preparation of this manuscript is gratefully acknowledged. The work in the author's laboratory was supported by grants from the NIH and The University of Texas M.D. Anderson Cancer Center.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • Of special interest
- •• Of outstanding interest

1 Heim S, Mitelman F: Cancer cytogenetics, edn 2. New York: Wiley Liss Inc., 1995.

2 Nowell PC: The clonal evolution of tumor cell populations. Science 1976, 194:23–28.

3 Tomlinson IP, Novelli MR, Bodmer WF: The mutation rate and cancer. Proc Natl Acad Sci USA 1996, 93:14800–14803.

4 •• Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. Nature 1998, 396:643–649.
An excellent review on the significance and possible mechanisms of genetic instability in cancer.

5 Heppner GH, Miller FR: The cellular basis of tumor progression. Int Rev Cytol 1998, 177:1–56.

6 Wolman SR: Chromosomal markers: signposts on the road to understanding neoplastic disease. Diag Cytopath 1998, 18:18–23.

7 Ross JS: DNA ploidy and cell cycle analysis in cancer diagnosis and prognosis. Oncology 1996, 10:867–890.

8 Magennis DP: Nuclear DNA in histological and cytological specimens: measurement and prognostic significance. Br J Biomed Sci 1997, 54:140–148.

9 Fletcher JA: Renal and bladder cancers. In: Human Cytogenetic Cancer Markers. Edited by Wolman SR, Sell S. Totowa, NJ: Humana Press; 1997:169–202.

10 Zhuang Z, Park WS, Pack S, Schmidt L, Vortmeyer AO, Pak E, et al.: Trisomy 7-harbouring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. Nat Genet 1998, 20:66–69.

11 Lindberg JO, Stenling RB, Rutegard JN: DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. Br J Surg 1999, 86:947–950.

12 Sasaki O, Kido K, Nagahama S: DNA ploidy, Ki-67 and p53 as indicators of lymph node metastasis in early gastric carcinoma. Anal Quant Cytol Histol 1999, 21:85–88.

13 Abad M, Ciudad J, Rincon MR, Silva I, Paz-Bouza JI, Lopez A, et al.: DNA aneuploidy by flow cytometry is an independent prognostic factor in gastric cancer. Anal Cell Path 1998, 16:223–231.

14 DeAngelis PM, Clausen OP, Schjolberg A, Stokke T: Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their

associations with tumour DNA ploidy, genotypes and phenotypes. Br J Cancer 1999, 80:526–535.

15 Malkhosyan S, Yasuda J, Scoto JL, Sekiya T, Yokota J, Perucho M: Molecular karyotype (amplotype) of metastatic colorectal cancer by unbiased arbitrarily primed PCR DNA fingerprinting. Proc Natl Acad Sci (USA) 1998, 95:10170–10175.

16 Ai H, Barrera JE, Pan Z, Meyers AD, Varella-Garcia M: Identification of individuals at high risk for head and neck carcinogenesis using chromosome aneuploidy detected by fluorescence in situ hybridization. Mut Res 1999, 439:223–232.

17 Barrera JE, Ai H, Pan Z, Meyers AD, Varella-Garcia M: Malignancy detection by molecular cytogenetics in clinically normal mucosa adjacent to head and neck tumors. Arch Otolaryngol Head Neck Surg 1998, 124:847–851.

18 Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al.: 17p (p53) allelic loss, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's oesophagus. Proc Natl Acad Sci USA 1996, 93:7081–7084.

19 Teodori L, Gohde W, Persiani M, Ferrario F, Tirindelli Danesi D, Scarpignato C, et al.: DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow up study on a cohort of patients. Cytometry 1998, 34:257–263.

20 Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, Paulson TG, et al.: Evolution of neoplastic cell lineages in Barrett oesophagus. Nat Genet 1999, 22:106–109.

21 Sturgis CD, Caraway NP, Johnston DA, Sherman SI, Kidd L, Katz RL: Image analysis of papillary thyroid carcinoma fine needle aspirates: significant association between aneuploidy and death from disease. Cancer 1999, 87:155–160.

22 Hemmer S, Wasenius VM, Knuutila S, Joensuu H, Franssila K: Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. Br J Cancer 1998, 78:1012–1017.

23 Mendelin J, Grayson M, Wallis T, Visscher DW: Analysis of chromosome aneuploidy in breast carcinoma progression by using fluorescence in situ hybridization. Lab Inv 1999, 79:387–393.

24 Bulten J, Poddighe PJ, Robben JC, Gemmink JH, deWilde PC, Hanselaar GAGJM: Interphase cytogenetic analysis of cervical intraepithelial neoplasia. Am J Pathol 1998, 152:495–503.

25 Krauter J, Ganser A, Bergmann L, Raghavachar A, Hoelzer D, Lübbert M, et al.: Association between structural and numerical chromosomal aberrations in acute myeloblastic leukemia: a study by RT-PCR and FISH in 447 patients with de novo AML. Ann Hematol 1999, 78:265–269.

26 Van Den Neste E, Louviaux I, Michaux JL, Delannoy A, Michaux L, Hagemeijer A, et al.: Myelodysplastic syndrome with monosomy 5 and/or 7 following therapy with 2-chloro-2'-deoxyadenosine. Br J Hematol 1999, 105:268–270.

27 Namba M, Mihara K, Fushimi K: Immortalization of human cells and its mechanisms. Crit Rev Oncog 1996, 7:19–31.

28 Li R, Yerganian G, Duesberg P, Kraemer A, Willer A, Rausch C, Hehlmann R: Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. Proc Natl Acad Sci USA 1997, 94:14506–14511.

29 Liu P, Zhang H, McLellan A, Vogel H, Bradley A: Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. Genetics 1998, 150:1155–1168.

30 Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC, Altieri DC: Control of apoptosis and mitotic spindle checkpoint survivin. Nature 1998, 396:580–584.

31 • Lengauer C, Kinzler KW, Vogelstein B: Genetic instability in colorectal cancers. Nature 1997, 386:623–627.
Demonstrates chromosomal instability in aneuploid colorectal tumor cells.

32 • Duesberg P, Rausch C, Rasnick D, Hehlmann R: Genetic instability of cancer cells is proportional to their degree of aneuploidy. Proc Natl Acad Sci USA 1998, 95:13692–13697.
Correlates aneuploidy and transformation in *in vitro* grown CHE cells.

33 Boveri T: Zur Frage der Entstehung maligner Tumoren. Jena, Verlag von Gustav Fischer, 1914.

34 • Bialy H: Aneuploidy and cancer: vintage wine in a new bottle? Nat Biotech 1998, 16:137–138.
Discusses the significance of aneuploidy and gene mutations in cancer.

35 Orr-Weaver TL, Weinberg RA: A checkpoint on the road to cancer. Nature 1998, 392:223–224.

36 • Pihan GA, Doxsey SJ: The mitotic machinery as a source of genetic instability in cancer. Semin Cancer Biol 1999, 9:289–302.
Describes various components and regulatory mechanisms of mitotic machinery and possible mechanisms of chromosome missegration in cancer.

37 Pihan GA, Purohit A, Wallace J, Knecht H, Ubda B, Queensberry P, Doxsey SJ: Centrosome defects and genetic instability in malignant tumors. Cancer Res 1998, 58:3974–3985.

38 Lingle WI, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. Proc Natl Acad Sci USA 1998, 95:2950–2955.

39 • Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al.: Tumor amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. Nat Genet 1998, 20:189–193.
Describes oncogenic property of centrosome associated STK15/aurora2 kinase and its involvement in aneuploidy induction.

40 • Bischoff JR, Anderson L, Shu Y, Morsie K, Ng I, Chan CS, et al.: A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. EMBO J 1998, 17:3052–3065.
Describes oncogenic property of STK15/aurora2 kinase and involvement in colorectal cancers.

41 Giet R, Uzbekov R, Cubizolles F, Le Guellec K, Prigent C: The xenopus laevis aurora related protein kinase pEq2 associates with and phosphorylates the Kinesin related protein X1Eq5. J Biol Chem 1999, 274:15005–15013.

42 Zimmerman W, Sparks C, Doxsey S: Amorphous no longer: the centrosome comes into focus. Curr Opin Cell Biol 1999, 11:122–128.

43 • Zou H, McGarry TJ, Bernal T, Kirschner MW: Identification of a vertebrate sister chromatid separation inhibitor involved in transformation and tumorigenesis. Science 1999, 285:418–421.
Demonstrates transforming and tumorigenic function of a gene inhibiting sister chromatid separation.

44 Hardwick KG: The spindle checkpoint. Trends Genet 1998, 14:1–4.

45 • Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JKV, et al.: Mutations of mitotic checkpoint genes in human cancers. Nature 1998, 392:300–303.
Describes mitotic checkpoint gene mutations in human colorectal cancers showing chromosome instability.

46 Jin DY, Spencer F, Jeang KT: Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. Cell 1998, 33:81–91.

47 Xu X, Weaver Z, Linke SP, Li C, Gotay J, Wang XW, et al.: Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform deficient cells. Mol Cell 1999, 3:389–395.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 273, No. 33, Issue of August 14, pp. 21161–21168, 1998
Printed in U.S.A.

Overexpression of a DEAD Box Protein (DDX1) in Neuroblastoma and Retinoblastoma Cell Lines*

(Received for publication, November 17, 1997, and in revised form, June 2, 1998)

Roseline Godbout‡, Mary Packer, and Wenjun Bie

From the Department of Oncology, Cross Cancer Institute and University of Alberta, 11560 University Ave., Edmonton, Alberta T6G1Z2, Canada

The DEAD box gene, *DDX1*, is a putative RNA helicase that is co-amplified with *MYCN* in a subset of retinoblastoma (RB) and neuroblastoma (NB) tumors and cell lines. Although gene amplification usually involves hundreds to thousands of kilobase pairs of DNA, a number of studies suggest that co-amplified genes are only overexpressed if they provide a selective advantage to the cells in which they are amplified. Here, we further characterize *DDX1* by identifying its putative transcription and translation initiation sites. We analyze DDX1 protein levels in *MYCN/DDX1*-amplified NB and RB cell lines using polyclonal antibodies specific to DDX1 and show that there is a good correlation with *DDX1* gene copy number, *DDX1* transcript levels, and DDX1 protein levels in all cell lines studied. DDX1 protein is found in both the nucleus and cytoplasm of *DDX1*-amplified lines but is localized primarily to the nucleus of nonamplified cells. Our results indicate that DDX1 may be involved in either the formation or progression of a subset of NB and RB tumors and suggest that DDX1 normally plays a role in the metabolism of RNAs located in the nucleus of the cell.

DEAD box proteins are a family of putative RNA helicases that are characterized by eight conserved amino acid motifs, one of which is the ATP hydrolysis motif containing the core amino acid sequence DEAD (Asp-Glu-Ala-Asp) (1–3). Over 40 members of the DEAD box family have been isolated from a variety of organisms including bacteria, yeast, insects, amphibians, mammals, and plants. The prototypic DEAD box protein is the translation initiation factor, eukaryotic initiation factor 4A, which, when combined with eukaryotic initiation factor 4B, unwinds double-stranded RNA (4). Other DEAD box proteins, such as p68, Vasa, and An3, can effectively and independently destabilize/unwind short RNA duplexes *in vitro* (5–7). Although some DEAD box proteins play general roles in cellular processes such as translation initiation (eukaryotic initiation factor 4A (4)), RNA splicing (PRP5, PRP28, and SPP81 in yeast (8–10)), and ribosomal assembly (SrmB in *Escherichia coli* (11)), the function of most DEAD box proteins remains unknown. Many of the DEAD box proteins found in higher eukaryotes are tissue- or stage-specific. For example, *PL10* mRNA is expressed only in the male germ line, and its product has been proposed to have a specific role in translational regulation during spermatogenesis (12). Vasa and ME31B are maternal proteins that may be involved in embryogenesis (13, 14). p68, found in dividing cells (15), is believed to be required for the formation of nucleoli and may also have a function in the regulation of cell growth and division (16, 17). Other DEAD box proteins are implicated in RNA degradation, mRNA stability, and RNA editing (18–20).

The human DEAD box protein gene *DDX1*[1] was identified by differential screening of a cDNA library enriched in transcripts present in the two RB cell lines Y79 and RB522A (21). The longest *DDX1* cDNA insert isolated from this library was 2.4 kb with an open reading frame from position 1 to 2201. All eight conserved motifs characteristic of DEAD box proteins are found in the predicted amino acid sequence of DDX1 as well as a region with homology to the heterogeneous nuclear ribonucleoprotein U, a protein believed to participate in the processing of heterogeneous nuclear RNA to mRNA (22, 23). The region of homology to heterogeneous nuclear ribonucleoprotein U spans 128 amino acids and is located between the first two conserved DEAD box protein motifs, 1a and 1b.

The proto-oncogene *MYCN* encodes a member of the MYC family of transcription factors that bind to an E box element (CACGTG) when dimerized with the MAX protein (24, 25). The *MYCN* gene is amplified and overexpressed in approximately one-third of all NB tumors (26, 27). Amplification of *MYCN* is associated with rapid tumor progression and a poor clinical prognosis (26, 27). MYCN overexpression is usually achieved by increasing gene copy number rather than by up-regulating basal expression of MYCN (27, 28). Because gene amplification involves hundreds to thousands of kilobase pairs of contiguous DNA (29–32), it is possible that co-amplification of a gene located in proximity to *MYCN* may contribute to the poor clinical prognosis of *MYCN*-amplified tumors. The *DDX1* gene maps to the same chromosomal band as *MYCN*, 2p24, and is located ~400 kb telomeric to the *MYCN* gene (33–36). All four *MYCN*-amplified RB tumor cell lines tested to date are amplified for *DDX1* (21),[2] while approximately two-thirds of NB cell lines and 38–68% of NB tumors are co-amplified for both genes (37–39). George *et al.* (39) found a significant decrease in the mean disease-free survival of patients with *DDX1/MYCN*-amplified NB tumors compared with *MYCN*-amplified tumors. Similarly, Squire *et al.* (38) observed a trend toward a worse clinical prognosis when both genes were amplified in the tumors of NB patients. To date, there have been no reports of a

* This work was supported by the National Cancer Institute of Canada with funds from the Canadian Cancer Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) X70649.

‡ To whom correspondence should be addressed: Dept. of Oncology, Cross Cancer Institute, 11560 University Ave., Edmonton, Alberta T6G 1Z2, Canada. Tel.: 403-432-8901; Fax: 403-432-8892.

[1] The abbreviations used are: DDX1, DEAD box 1; NB, neuroblastoma; RB, retinoblastoma; RACE, rapid amplification of cDNA ends; PAGE, polyacrylamide gel electrophoresis; nt, nucleotide(s); MOPS, 4-morpholinepropanesulfonic acid; bp, base pair(s); kb, kilobase(s) or kilobase pair(s).

[2] R. Godbout, unpublished results.

tumor amplified only for *DDX1*, and the role that this gene plays in cancer formation and progression is not known.

Because of the high rate of rearrangements in amplified DNA (31, 40), it is unlikely that a gene located ~400 kb from the *MYCN* gene will be consistently amplified as an intact unit unless its product provides a growth advantage to the cell. Based on Southern blot analysis, the *DDX1* gene extends over more than 30 kb, and there are no gross rearrangements of this gene in *DDX1*-amplified tumors (21, 38). Furthermore, there is a good correlation between *DDX1* transcript levels and gene copy number in the tumors analyzed to date. However, we need to show that DDX1 protein is overexpressed in *DDX1*-amplified tumors if we are to entertain the possibility that this protein plays a role in the tumorigenic process. Here, we isolate and characterize the 5′-end of *DDX1* mRNA and extend the *DDX1* cDNA sequence by ~300 nt. We identify the predicted initiation codon of DDX1 and generate antisera that specifically recognize DDX1 protein. We analyze levels of DDX1 protein in both *DDX1*-amplified and nonamplified RB and NB tumors and study the subcellular location of this protein in the cell.

MATERIALS AND METHODS

Library Screening—A human fetal brain cDNA library (Stratagene) was screened using a 320-bp DNA fragment from the 5′-end of the 2.4-kb *DDX1* cDNA previously described (23). Phagemids containing positive inserts were excised from λ ZAP II following the supplier's directions. The ends of the cDNA inserts were sequenced using the dideoxynucleotide chain termination method with T7 DNA polymerase (Amersham Pharmacia Biotech).

A human placenta genomic library (CLONTECH) was screened with the 5′-end of *DDX1* cDNA. Positive plaques were purified, and the genomic DNA was analyzed using restriction enzymes and Southern blotting. *Eco*RI-digested DNA fragments from these clones were subcloned into pBluescript and digested with exonuclease III and mung bean nuclease to obtain sequentially deleted clones. The exon/intron map of the 5′ portion of the *DDX1* gene was obtained by comparing the sequence of *DDX1* cDNA with that of the genomic DNA.

Rapid Amplification of cDNA Ends (RACE)—We used the AmpliFINDER RACE kit (CLONTECH) to extend the 5′-end of *DDX1* cDNA. Briefly, two μg of poly(A)$^+$ RNA isolated from RB522A was reverse transcribed at 52 °C using either primer P1 or P3 (Fig. 1*A*). The RNA template was hydrolyzed, and excess primer was removed. A single-stranded AmpliFINDER anchor containing an *Eco*RI site was ligated to the 3′-end of the cDNA using T4 RNA ligase. The cDNA was amplified using either primer P2 or P4 (Fig. 1*A*) and AmpliFINDER anchor primer. RACE products were cloned into pBluescript.

Primer Extension—Poly(A)$^+$ RNAs were isolated from RB and NB cell lines as described previously (21, 38). The 21-nt primers 5′-TTCGTTCTGGGCACCATGTGT-3′ (primer P4 in Fig. 1*A*) and 5′-TGGGACCTAGGGCTTCTGGAC-3′ (primer P3 in Fig. 1*A*) were end-labeled with [γ-^{32}P]ATP (3000 Ci/mmol; Mandel Scientific) and T4 polynucleotide kinase. Each of the labeled primers was annealed to 2 μg of poly(A)$^+$ RNA at 45 °C for 90 min, and the cDNA was extended at 42 °C for 60 min using avian myeloblastosis virus reverse transcriptase (Promega). The primer extension products were heat-denatured and run on a 8% polyacrylamide gel containing 7 M urea in 1× TBE buffer. A G + A sequencing ladder served as the size standard.

S1 Nuclease Protection Assay—The S1 nuclease protection assay to map the transcription initiation site of *DDX1* was performed as described by Favaloro *et al.* (41). The DNA probe was prepared by digesting genomic DNA spanning the upstream region of *DDX1* and exon 1 with *Ava*I, labeling the ends with [γ-^{32}P]ATP (3000 Ci/mmol) and polynucleotide kinase, and removing the label from one of the ends by digesting the DNA with *Sph*I (Fig. 4). The RNA samples were resuspended in a hybridization mixture containing 80% formamide, 40 mM PIPES, 400 mM NaCl, 1 mM EDTA, and the heat-denatured *Sph*I–*Ava*I probe labeled at the *Ava*I site. The samples were incubated at 45 °C for 16 h and digested with 3000 units/ml S1 nuclease (Boehringer Mannheim) for 60 min at 37 °C. The samples were precipitated with ethanol; resuspended in 80% formaldehyde, TBE buffer, 0.1% bromphenol blue, xylene cyanol; denatured at 90 °C for 2 min; and electrophoresed in a 7 M urea, 8% polyacrylamide gel in TBE buffer.

Northern and Southern Blot Analysis—Poly(A)$^+$ RNAs were isolated from RB and NB cell lines as described previously (21, 38). Two μg of poly(A)$^+$ RNA/lane were electrophoresed in a 6% formaldehyde, 1.5% agarose gel in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0) and transferred to nitrocellulose filter in 3 M sodium chloride, 0.3 M sodium citrate. The filters were hybridized to the following DNA probes, ^{32}P-labeled by nick translation: (i) a 1.6-kb *Eco*RI insert from *DDX1* cDNA clone 1042 (21), (ii) a 260-bp cDNA fragment spanning the 3′-end of *DDX1* exon 1 as well as exons 2 and 3, (iii) a 160-bp fragment derived from the 5′-end of *DDX1* exon 1, and (iv) α-actin cDNA to control for lane to lane variation in RNA levels. Filters were hybridized and washed under high stringency. Southern blot analysis was as described previously (21).

Preparation of Anti-DDX1 Antiserum—To prepare antiserum to the C terminus of the DDX1 protein, we inserted a 1.8-kb *Eco*RI fragment from bp 848 to 2668 of *DDX1* cDNA (Fig. 1*B*) into *Eco*RI-digested pMAL-c2 expression vector (New England Biolabs). DH5α cells transformed with this vector were grown to mid-log phase and induced with 0.1 mM isopropyl-1-thio-β-D-thiogalactoside. The cells were harvested 3–4 h postinduction and lysed by sonication. Soluble maltose binding protein-DDX1 fusion protein was affinity-purified using amylose resin, and the maltose-binding protein was cleaved with factor Xa. The DDX1 protein was purified on a SDS-PAGE gel, electroeluted, and concentrated. Approximately 100 μg of protein was injected into rabbits at 4–6-week intervals. For the initial injection, the protein was dispersed in complete Freund's adjuvant (Sigma), while subsequent injections were prepared in Freund's incomplete adjuvant. Blood was collected from each rabbit 10 days after injection, and the specificity of the antiserum was tested using cell extracts from RB522A. To prepare antiserum to the N terminus of DDX1 protein, a *DDX1* cDNA fragment from bp 268 to 851 (Fig. 1*B*) was inserted into pGEX-4T2 (Amersham Pharmacia Biotech). The recombinant protein produced from this construct contains the first 186 amino acids of the predicted DDX1 sequence. Soluble glutathione *S*-transferase-DDX1 fusion protein was purified with glutathione-Sepharose 4B (Amersham Pharmacia Biotech). The glutathione *S*-transferase component of the fusion protein was cleaved with thrombin.

Subcellular Fractionations and Western Blot Analysis—We used two different procedures for subcellular fractionations. First, we isolated nuclear and S100 (soluble cytoplasmic) fractions from RB522A, IMR-32, Y79, RB(E)-2, HeLa, and HL60 using the procedure of Dignam (42). On average, we obtained 5–6 times more protein in the cytosolic fractions than in the nuclear fractions. Second, 10^8 RB522A cells were lysed and fractionated into S4 (soluble cytoplasmic components), P2 (heavy mitochondria, plasma membrane fragments), P3 (mitochondria, lysozymes, peroxisomes, and Golgi membranes), and P4 fractions (membrane vesicles from rough and smooth endoplasmic reticulum, Golgi, and plasma membrane) by differential centrifugation (43). We obtained 8 mg of protein in the S4 fraction, 1 mg in P2, 0.5 mg in P3, and 2 mg in P4 fraction. The procedures related to the immunoelectron microscopy have been previously described (44).

For Western blot analysis, proteins were electrophoresed in polyacrylamide-SDS gels and electroblotted onto nitrocellulose using the standard protocol for protein transfer described by Schleicher and Schuell. The filters were incubated with a 1:5000 dilution of DDX1 antiserum, a 1:200 dilution of anti-MYCN monoclonal antibody (Boehringer Mannheim), or a 1:200 dilution of anti-actin (Santa Cruz Biotechnology, Inc., Santa Cruz, CA). For the colorimetric analysis, antigen-antibody interactions were visualized using either alkaline phosphatase-linked goat anti-rabbit IgG (for DDX1) or goat anti-mouse IgG (for MYCN) at a 1:3000 dilution. For the ECL Western blotting analysis (Amersham Pharmacia Biotech), we used a 1:100,000 dilution of peroxidase-linked secondary anti-rabbit IgG antibody (for DDX1) or secondary anti-goat IgG antibody (Jackson ImmunoResearch Laboratories).

RESULTS

Identification of the 5′-End of the DDX1 Transcript—We have previously reported the sequence of *DDX1* cDNA isolated from an RB cDNA library (21, 23). This 2.4-kb *DDX1* cDNA contains an open reading frame spanning positions 1–2201 with a methionine encoded by the first three nucleotides (Fig. 1*A*). There is a polyadenylation signal and poly(A) tail in the 3′-untranslated region, indicating that the sequence is complete at the 3′-end. Manohar *et al.* (37) have also isolated *DDX1* cDNA from the NB cell line LA-N-5. Their cDNA extended the 5′-end of our sequence by 42 bp and included an additional in frame methionine (*double underlined* in Fig. 1*A*). The possibil-

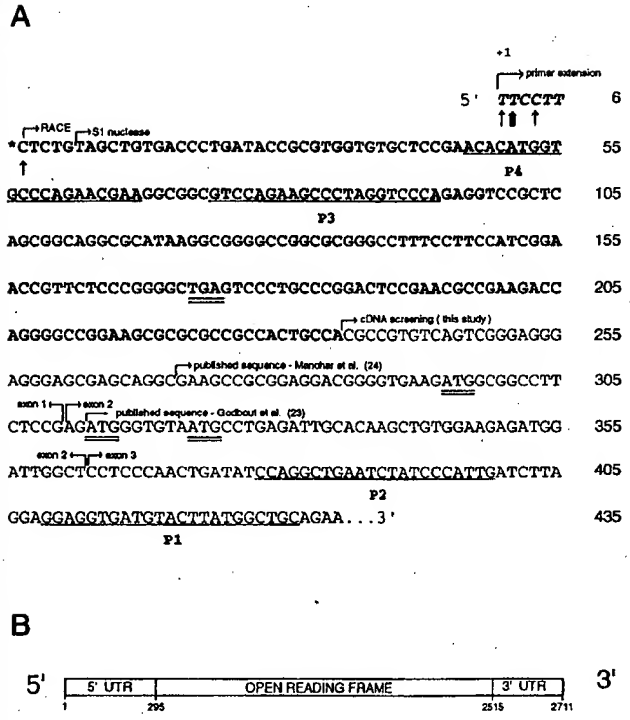

FIG. 1. **Partial sequence and structure of *DDX1* cDNA.** *A*, the sequence of the 5′-end of *DDX1* cDNA. The sequence in *boldface type* starting at the *asterisk* was obtained using the RACE strategy. The additional 6 bp in *italic boldface type* at the 5′-end of the cDNA are predicted based on the known *DDX1* genomic sequence and primer extension analysis. P1, P2, P3, and P4 are primers used in the RACE experiments (the complementary sequence was used in each case). Primers P3 and P4 were also used for the primer extension analysis. Three in frame methionine codons are indicated by the *double underline*. An in frame stop codon is indicated by the *boldface double underline*. The three major transcription initiation sites identified by primer extension are indicated by the *single arrows*, while a minor site is represented by the *broad arrow*. The predicted *DDX1* transcription initiation sites obtained by RACE, S1 nuclease, and primer extension are indicated as well as the 5′-ends of *DDX1* cDNA sequences obtained by screening cDNA libraries. The sequences transcribed from exons 1, 2, and 3 are also shown. *B*, the structure of the 2711-bp *DDX1* cDNA is shown with an open reading frame from position 295 to 2515.

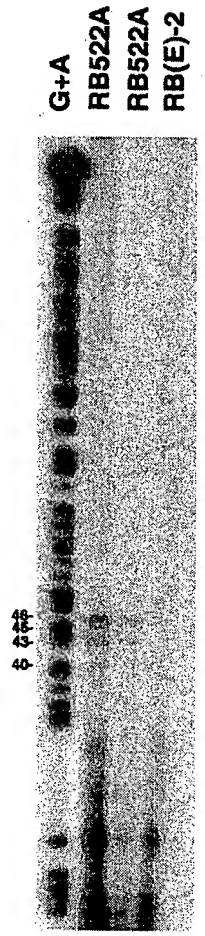

FIG. 2. **Identification of the 5′-end of the *DDX1* transcript by primer extension.** Radioactively labeled primer P4 was annealed to 2 μg of poly(A)$^+$ RNA from RB522A (*lane 1*), 1 μg of poly(A)$^+$ RNA from RB522A (*lane 2*), and 2 μg of poly(A)$^+$ RNA from RB(E)-2 cells (*lane 3*), and extended using reverse transcriptase. The products were run on an 8% denaturing polyacrylamide gel with a G + A sequencing ladder as size marker. The primer extension products are indicated on the *left*. The sizes of the products (in nt) are presented as the distance from primer P4.

ity of additional in frame methionines located further upstream could not be excluded, because there were no predicted stop codons in the upstream region of the cDNA.

Northern blot analysis indicated a *DDX1* transcript size of ~2800 nt, suggesting that the *DDX1* cDNAs isolated to date were lacking ~300–350 bp of 5′ sequence. We have used different approaches to identify the transcription start site of *DDX1*. First, we exhaustively screened a commercial fetal brain cDNA library with the 5′-end of *DDX1* cDNA. Although numerous clones were analyzed, only one extended the sequence (by 35 bp) beyond that published by Manohar *et al.* (37) (Fig. 1*A*).

We next used the RACE procedure in an attempt to isolate additional 5′ sequence. The nested primers used to amplify the 5′-end of the *DDX1* transcript are labeled as primers P1 and P2 in Fig. 1*A* and are located downstream of the three in frame methionines (*double underlined* in Fig. 1*A*). Poly(A)$^+$ RNA from RB522A was reverse transcribed at 52 °C using primer P1, and the reverse transcribed cDNA was amplified using the nested primer P2 and the 5′-RACE primer. Using this approach, we generated a product that was 230 bp longer than any of the cDNAs obtained by screening libraries (Fig. 1*A*). Sequencing of this 230-bp cDNA revealed an in frame stop codon (*boldface double underline* in Fig. 1*A*) located 123 bp upstream of the predicted translation initiation site. We then prepared primers P3 and P4, located near the 5′-end of the RACE cDNA (Fig. 1*A*) and repeated the RACE procedure to see if additional 5′ sequences could be obtained. The resulting RACE products did not extend the *DDX1* cDNA sequence further.

The location of the *DDX1* transcription initiation site was verified by primer extension. Poly(A)$^+$ RNA was prepared from the following two cell lines: *DDX1*-amplified RB cell line RB522A and a nonamplified RB cell line RB(E)-2. RB522A has elevated levels of *DDX1* mRNA, while RB(E)-2 has at least 20-fold lower levels of this transcript. Three products of 40, 43, and 46 nt (with a weak signal at 45 nt) were detected in RB522A using primer P4 (Figs. 1*A* and 2). The 40-nt product corresponded exactly with the 5′-end of the RACE-derived cDNA while the 43- and 46-nt products extended the predicted size of the *DDX1* transcript by 3 and 6 nt, respectively. None of these products were observed in RB(E)-2. Bands of identical sizes to those obtained with RB522A mRNA were also observed in the *DDX1*-amplified NB cell line BE(2)-C but not in the *DDX1*-amplified NB cell line IMR-32 (data not shown). The same predicted *DDX1* transcription initiation site was identified with primer P3 except that the bands were of weaker

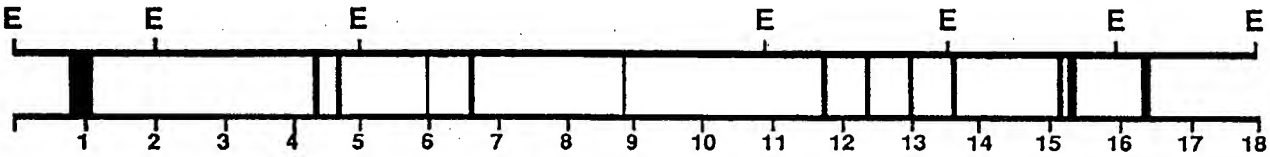

FIG. 3. **Genomic map of the 5′-end of *DDX1*.** The exons are represented by the *black boxes*, and distances are in kilobase pairs. The locations of *Eco*RI (*E*) sites are indicated.

intensity (data not shown). We have designated the transcription start site identified by primer extension as +1 (Fig. 1*A*).

The sequence of the 6 nt extending beyond the RACE cDNA was obtained by comparison of the cDNA sequence with that of *DDX1* genomic DNA. Bacteriophages containing *DDX1* genomic DNA were isolated by screening a human placenta library with 5′ *DDX1* cDNA. Eighteen kb of DNA were sequenced from two bacteriophages with overlapping *DDX1* genomic DNA. Thirteen exons were identified within this 18-kb region (Fig. 3) corresponding to cDNA sequences from position 1 to 1249. The 310-bp exon 1 was by far the longest of the 13 exons sequenced, corresponding to the entire 5′-untranslated region of *DDX1* as well as the first in frame methionine. The sequences transcribed from exons 1, 2, and 3 are indicated in Fig. 1*A*.

Knowledge of the genomic structure of *DDX1* allowed us to use the S1 protection assay, a technique that is independent of reverse transcriptase, to further define the 5′-end of the *DDX1* transcript. Poly(A)$^+$ RNAs from six *DDX1*-amplified lines (RB lines: Y79 and RB522A; NB lines: BE(2)-C, IMR-32, LA-N-1, and LA-N-5) and six nonamplified lines (RB lines: RB(E)-2 and RB412; NB lines, GOTO, NB-1, NUB-7, and SK-N-MC) were hybridized to a DNA probe that extended from position −745 in the 5′-flanking *DDX1* DNA to position +164 in exon 1. This DNA probe was labeled at position +164 as indicated in Fig. 4. Nonhybridized DNA was digested with S1 nuclease, and the sizes of the protected fragments were analyzed on a denaturing polyacrylamide gel. Bands of 150–153 nt were observed in *lane 2* (RB522A), *lane 5* (BE(2)-C), and *lane 8* (LA-N-1) with bands of much weaker intensity in *lane 7* (IMR-32) (Fig. 4). Specific bands were not detected in either *DDX1*-amplified Y79 and LA-N-5 or the nonamplified lines. Although the sizes of the S1 protected bands in RB522A, BE(2)-C, and LA-N-1 were 5 and 11 nt shorter than predicted based on RACE and primer extensions, respectively, there was general agreement with all three techniques regarding the location of the *DDX1* transcription initiation site (Fig. 1*A*). The smaller S1 nuclease protected products could have arisen as the result of S1 digestion of the 5′-end of the RNA:DNA heteroduplex because of its relatively high rU:dA content (45).

Identification of the same transcription initiation site in three *DDX1*-amplified lines suggests that this represents the bona fide start site of *DDX1* transcription. However, it was not clear why this start site was either very weak or not detected in three other amplified lines. To determine whether the 5′-end of exon 1 is transcribed in all *DDX1*-amplified lines, we carried out a direct analysis of the 5′-end of the *DDX1* transcript by Northern blotting. Two probes were used for this analysis: the 5′ probe contained a 160-bp fragment from bp 1 to 160 (5′-half of exon 1), and the 3′ probe contained a 260-bp fragment from bp 160 to 420 (3′-half of exon 1 as well exons 2 and 3) (Fig. 1*A*). With the 3′ probe, we obtained bands of similar size and intensity in four *DDX1*-amplified lines (RB522A, BE(2)-C, IMR-32, and LA-N-5). Band intensity was somewhat weaker in Y79 and stronger in LA-N-1 in comparison with the other lines (Fig. 5). No signal was detected in the non-*DDX1*-amplified line RB412. With the 5′ probe, a relatively strong signal was observed in RB522A, BE(2)-C, and LA-N-1, while a considerably weaker but readily apparent signal was detected in Y79, IMR-32, and LA-N-5. The signal obtained with actin indicates that, with the exception of LA-N-1, similar amounts of RNA were loaded in each lane and that the RNA was not degraded. These results indicate that at least a portion of the 160-bp 5′-end of exon 1 is transcribed in all *DDX1*-amplified lines.

Based on primer extension, S1 nuclease protection assay, Northern blot analysis and the sequencing of the RACE products, we conclude that the *DDX1* transcript is 2.7 kb with an open reading frame spanning nucleotides 295–2515 encoding a predicted protein of 740 amino acids with an estimated molecular weight of 82.4 (Fig. 1*B*). An in frame stop codon is located 123 nt upstream of the predicted translation initiation site, at positions 172–174. The first in frame methionine following the stop codon is in agreement with the Kozak consensus sequence (46). Furthermore, the predicted start methionine codon for human DDX1 corresponds perfectly with that of *Drosophila* DDX1 (47). A stop codon is located 15 nt upstream of the initiation codon in *Drosophila* DDX1.

Analysis of DDX1 Protein Levels in Neuroblastoma and Retinoblastoma—We and others have previously shown that there is a good correlation between gene copy number and RNA levels in *DDX1*-amplified RB and NB cell lines (37, 38). To determine whether the correlation extends to DDX1 protein levels, we prepared antiserum to two nonoverlapping recombinant DDX1 proteins. First, we prepared a C terminus recombinant protein construct by inserting a 1.8-kb *Eco*RI fragment from bp 848 to 2668 (amino acids 185–740) (Fig. 1*B*) into the pMAL-c2 expression vector. Recombinant protein expression was induced with isopropyl-1-thio-β-D-thiogalactoside, and the 110-kDa maltose-binding protein-DDX1 fusion product was purified by affinity chromatography using amylose resin, followed by electrophoresis on a SDS-PAGE gel after cleaving the maltose-binding protein fusion partner with factor Xa. Second, we prepared an N terminus construct by ligating a DNA fragment from bp 268 to 851 (amino acids 1–186) into pGEX-4T2. The 50-kDa glutathione *S*-transferase-DDX1 fusion protein was purified by affinity chromatography on a glutathione column. This N terminus fusion protein contains only the first of the eight conserved motifs found in all DEAD box proteins, while the C terminus fusion protein includes the remaining seven motifs.

We measured DDX1 protein levels in total cell extracts of three RB and 10 NB cell lines. Using antiserum to the N terminus fusion protein, we observed a strong signal in all *DDX1*-amplified cell lines: the RB cell lines Y79 (*lane 1*) and RB522A (*lane 2*) and the NB cell lines BE(2)-C (*lane 4*), IMR-32 (*lane 6*), LA-N-1 (*lane 8*), and LA-N-5 (*lane 9*) (Fig. 6). Two bands were observed in the majority of extracts. Of the amplified lines, Y79 produced the weakest signal, with the most intense signal observed in LA-N-1. There was an excellent correlation with DDX1 protein and mRNA levels in these cell lines, with lower levels of DDX1 mRNA observed in Y79 and higher levels in LA-N-1 (Fig. 7*A*). As shown in Fig. 7*B*, this correlation extended to *DDX1* gene copy number. No gross DNA rearrangements were seen in the *DDX1*-amplified lines; however, three small bands of altered size were observed in the RB412 lane. Although the nature of the DNA alteration is not known, it is noteworthy that *DDX1* transcript levels in RB412

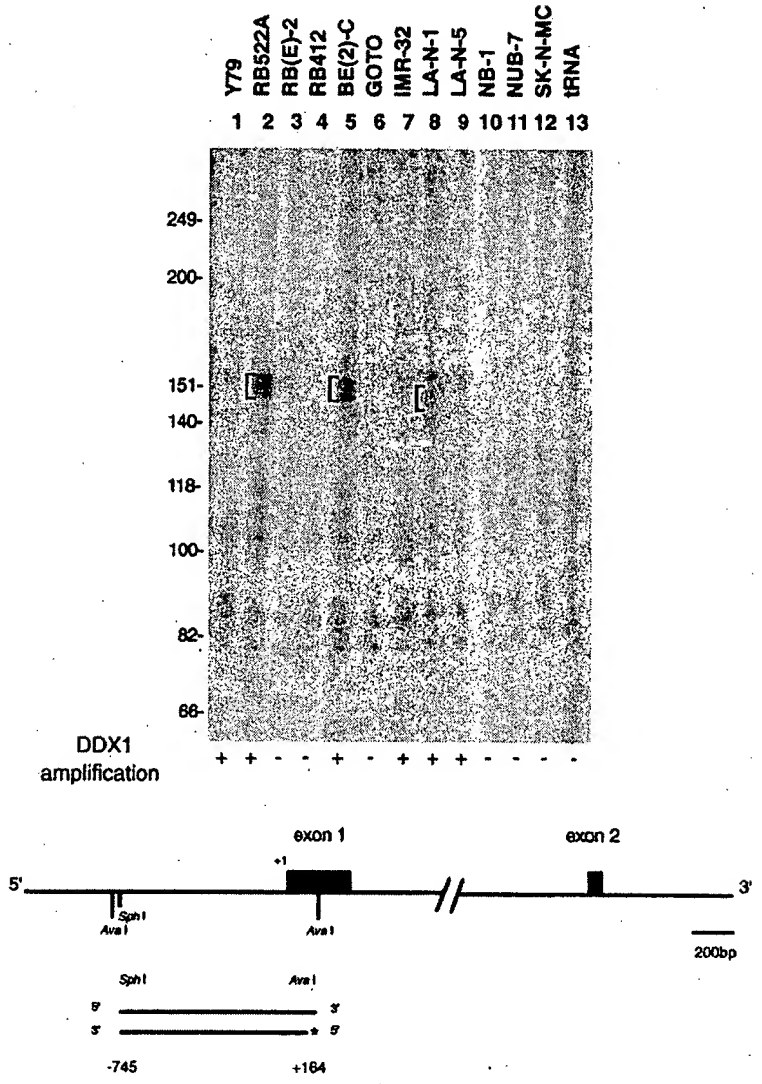

FIG. 4. **S1 nuclease mapping of the 5′-end of the *DDX1* transcript.** Two μg of poly(A)$^+$ RNA from four RB lines (*DDX1*-amplified Y79 and RB522A and nonamplified RB(E)-2 and RB412), eight NB lines (*DDX1*-amplified BE(2)-C, IMR-32, LA-N-1, and LA-N-5 and nonamplified GOTO, NB-1, NUB-7, and SK-N-MC), and tRNA as a negative control were hybridized to a *Sph*I–*Ava*I fragment labeled at the *Ava*I site with [γ-^{32}P]ATP and polynucleotide kinase. Bands of 150–153 nt are shown in *lanes 2* (RB522A), *5* (BE(2)-C), and *8* (LA-N-1) with much weaker bands in *lane 7* (IMR-32). A map of the probe indicating the transcription initiation site identified by primer extension (+*1*), the labeling site (*), and exons 1 and 2, is shown at the *bottom*.

are extremely low (Fig. 7*A*) and that the top DDX1 protein band in RB412 cell extracts is smaller in size than the top band from the other cell extracts (Fig. 6).

Two DDX1 protein bands were present in most of the lanes in Fig. 6. The same two bands were detected with antiserum to the C terminus of the DDX1 protein, as well as a third band at ~60 kDa (data not shown). There was no variation in the intensity of the 60-kDa band in *DDX1*-amplified and nonamplified cell extracts. The 60-kDa band probably represents another member of the DEAD box protein family, because the C terminus DDX1 protein used to prepare this antiserum contained seven of the eight conserved motifs found in all DEAD box proteins. To obtain an estimate of the size of the two DDX1 bands, we ran cellular extracts from RB522A on a 7% SDS-PAGE gel with the BenchMark protein ladder (Life Technologies, Inc.). The size of the DDX1 protein was determined using the Alpha Imager 2000 documentation and analysis system for molecular weight calculation. Based on this analysis, the estimated molecular mass of the top band is 89.5 kDa, while that of the bottom band is 83.5 kDa. The 84-kDa band may represent the unmodified product encoded by the DDX1 transcript (capable of encoding a protein with a predicted molecular mass of 82.4 kDa), while the top band may represent post-translational modification of DDX1 protein (*e.g.* phosphorylation). Another possibility is that the top band represents intact DDX1

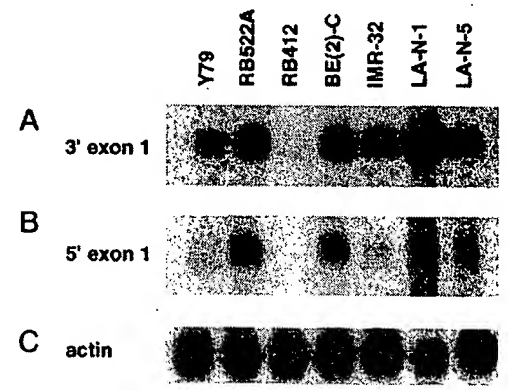

FIG. 5. **Northern blot analysis of the 5′-end of the *DDX1* transcript.** Two μg of poly(A)$^+$ RNA isolated from *DDX1*-amplified Y79, RB522A, BE(2)-C, IMR-32, LA-N-1, and LA-N-5 and nonamplified RB412 were electrophoresed in a 1.5% agarose-formaldehyde gel. The RNA was transferred to a nitrocellulose filter and sequentially hybridized with a 260-bp fragment from *DDX1* cDNA from bp +160 to +420 (3′-end of exon 1 as well as exons 2 and 3) (*A*), a 160-bp fragment from *DDX1* cDNA from bp +1 to +160 (5′-end of exon 1) (*B*), and actin cDNA (*C*). The DNA was labeled with [^{32}P]dCTP by nick translation. The blots were hybridized and washed under high stringency.

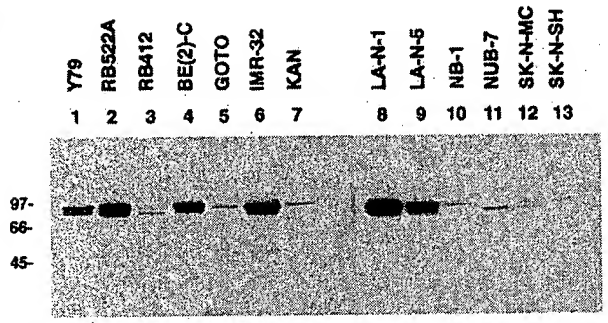

FIG. 6. **DDX1 protein expression in RB and NB cell lines.** Western blots were prepared using total cellular extracts from three RB (Y79, RB522A, and RB412) and 10 NB cell lines (BE(2)-C, GOTO, IMR-32, KAN, LA-N-1, LA-N-5, NB-1, NUB-7, SK-N-MC, and SK-N-SH). The lines that are amplified for the *DDX1* gene are Y79, RB522A, BE(2)-C, IMR-32, LA-N-1, and LA-N-5. Twenty μg of protein were loaded in each lane and electrophoresed in a 10% SDS-PAGE gel. DDX1 was detected using a 1:5000 dilution of the antiserum to the amino terminus of DDX1 protein. Size markers in kilodaltons are indicated on the *side*.

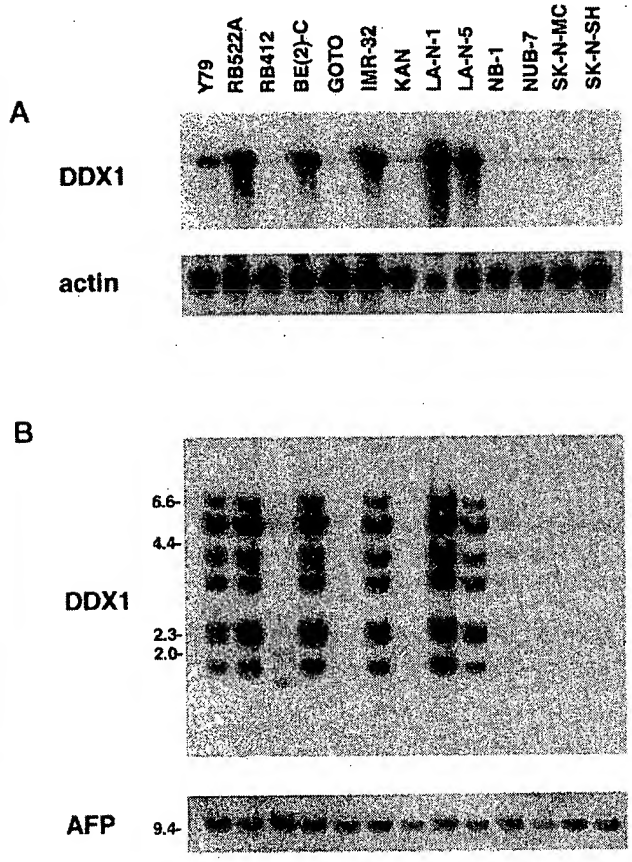

FIG. 7. **Northern and Southern blot analyses of *DDX1* in RB and NB cell lines.** *A*, 2 μg of poly(A)$^+$ RNA were loaded in each lane, electrophoresed in a 1.5% agarose-formaldehyde gel, and transferred to a nitrocellulose filter. The filter was first hybridized to a ^{32}P-labeled 1.6-kb DDX1 cDNA (clone 1042) (21), stripped, and rehybridized to actin DNA. *B*, 10 μg of genomic DNA from each of the indicated cell lines were digested with *Eco*RI, electrophoresed in a 1% agarose gel, and transferred to a nitrocellulose filter. The filter was hybridized to ^{32}P-labeled clone 1042 *DDX1* cDNA, stripped, and reprobed with labeled α-fetoprotein cDNA. Markers (in kilobase pairs) are indicated on the *side*.

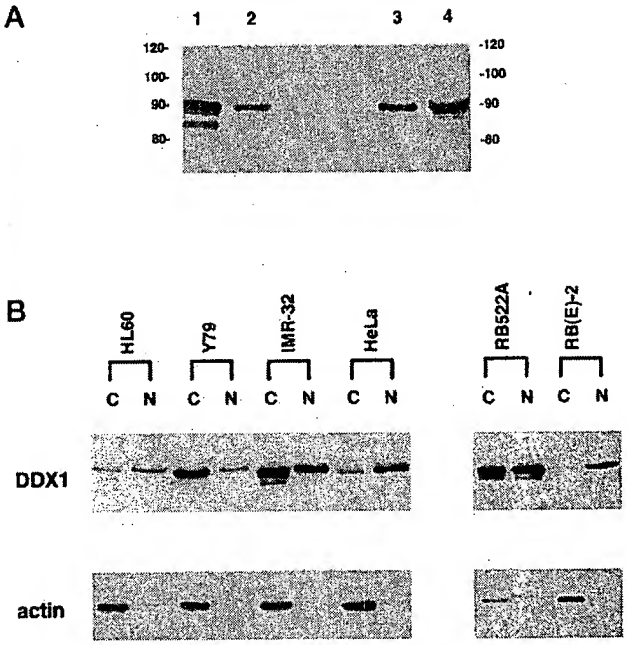

FIG. 8. **Distribution of DDX1 in the nucleus and cytoplasm.** *A*, cytosolic and nuclear extracts were prepared from RB522A and electrophoresed in a 7% SDS-PAGE gel. Cytosolic extracts were loaded in *lanes 1* (20 μg of protein) and *2* (10 μg), while nuclear extracts were loaded in *lanes 3* (10 μg) and *4* (20 μg). DDX1 was visualized using a 1:5000 dilution of the antiserum to the N terminus. The BenchMark protein ladder size markers (kilodaltons) are indicated on the *left*. *B*, cytosolic and nuclear extracts were prepared from HL60, Y79, IMR-32, HeLa, RB522A, and RB(E)-2 and electrophoresed in an 8% SDS-PAGE gel. Twenty μg of proteins were loaded in each *lane* marked *C* (cytosolic) and *N* (nuclear). DDX1 was visualized using a 1:5000 dilution of the antiserum to the N terminus. Actin levels were analyzed using a 1:200 dilution of anti-actin antibody (Santa Cruz Biotechnology).

and the lower band is a specific truncated or degradation product of DDX1. Yet a third possibility is that the two bands represent the products of differentially spliced transcripts or different translation initiation sites. However, the lack of any obvious differences in *DDX1* transcript sizes in the three RB and 10 NB lines analyzed in Fig. 7*A* does not support the latter possibility (*e.g.* compare the *DDX1* transcript size in NUB-7 (which produces the lower DDX1 protein band) and in NB-1 (which produces the higher DDX1 protein band)).

Subcellular Localization of DDX1 Protein—DEAD box proteins have been implicated in a variety of cellular functions including RNA splicing in the nucleus, translation initiation in the cytoplasm, and ribosome assembly in the nucleolus. To obtain an indication of the possible role of DDX1, we studied its subcellular location. Nuclear and cytosolic extracts were prepared from *DDX1*-amplified RB522A and run on a 7% SDS-PAGE gel. Although there was more DDX1 protein in the cytosol than in the nucleus on a per cell basis, the proportion of DDX1 protein relative to total protein was similar in both cellular compartments (Fig. 8*A*). Both the 90- and 84-kDa bands were present in cytosol and nuclear extracts, although the bottom band was more readily apparent in the cytosol. By running the gel for an extended period of time (twice as long as usual), we were able to detect an additional weak band at ~88 kDa in both nuclear and cytosolic extracts.

To determine whether DDX1 consistently localizes to both the cytoplasm and nucleus, we prepared cytosol and nuclear extracts from two additional *DDX1*-amplified lines, Y79 and IMR-32, as well as from nonamplified RB(E)-2, HL60, and HeLa. DDX1 protein was found in both the nucleus and cytoplasm of IMR-32, primarily in the cytoplasm of Y79, and mainly in the nucleus of the three nonamplified lines (Fig. 8*B*). In addition, DDX1 was almost exclusively found in nuclear

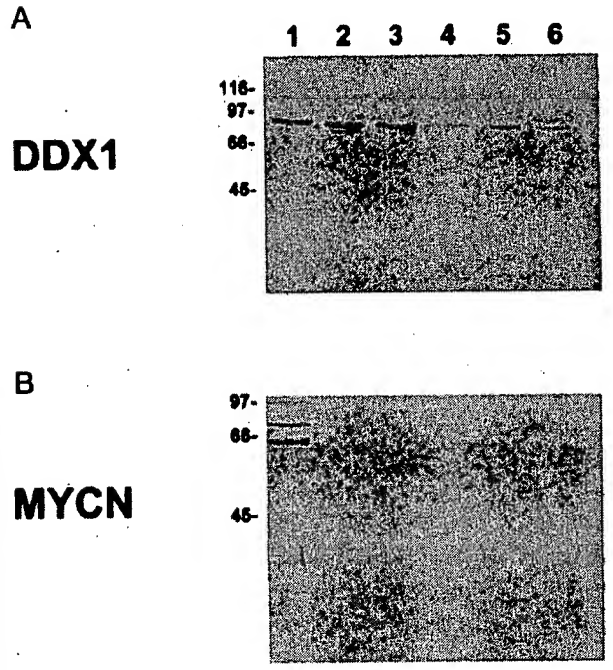

FIG. 9. **Subcellular location of DDX1 protein.** RB522A cells were fractionated into nuclear (*lane 1*), S100 and S4 cytosol (*lanes 2* and *3*), P2 membrane (*lane 4*), P3 membrane (*lane 5*), and P4 membrane (*lane 6*) fractions. Twenty μg of protein were loaded in each lane and run on a 10% SDS-PAGE gel. *A*, DDX1 protein was detected using a 1:5000 dilution of the antiserum to the N terminus of DDX1. *B*, MYCN protein was detected using a commercially available antibody at a 1:200 dilution. Size markers (kilodaltons) are indicated on the *side*.

extracts prepared from normal GM38 fibroblasts (data not shown). We used anti-actin antibody to ensure that our nuclear and cytosolic extracts were not cross-contaminated (Fig. 8*B*).

We next carried out a more detailed analysis of DDX1 subcellular location using two different approaches: (i) fractionation of cellular components into nuclei; S100 or S4 cytosol (containing soluble cytoplasmic components, including 40 S ribosomes); P2 (heavy mitochondria, plasma membrane fragments plus material trapped by these membranes); P3 (mitochondria, lysosomes, peroxisomes, Golgi membranes, some rough endoplasmic reticulum); and P4 (microsomes from smooth and rough endoplasmic reticulum, Golgi and plasma membranes) (43); and (ii) immunogold electron microscopy. The *DDX1*-amplified RB522A cell line was used for both experiments. The fractionation procedures indicate that DDX1 is mainly in the nucleus and in the cytosol (S4 and S100 fractions) of RB522A cells (Fig. 9*A*). As a control, we used anti-human MYCN antibody to determine the location of MYCN (also amplified in RB522A) in our subcellular fractions. As shown in Fig. 9*B*, MYCN was primarily found in the nucleus, as one would expect of a transcription factor.

For the electron microscopy analysis, antiserum to the N terminus of DDX1 was coupled to protein A gold particles, and the distribution of DDX1 was examined in RB522A cells fixed in paraformaldehyde and glutaraldehyde. DDX1 was present in both the cytoplasm and nucleus (data not shown). There was no association with either cell organelles or with nuclear or plasma membranes.

DISCUSSION

There are presently few clues as to the function of DDX1 in normal and cancer cells. Our earlier data indicate that *DDX1* mRNA is present at higher levels in fetal tissues of neural origin (retina and brain) compared with other fetal tissues (21). There may therefore be a requirement for elevated levels of this putative RNA helicase for the efficient production or processing of neural specific transcripts. A role in cancer formation or progression is an intriguing possibility, because overexpression of an RNA unwinding protein could affect the secondary structure of RNAs in such a way as to alter the expression of specific proteins in tumor cells. *DDX1* is co-amplified with *MYCN* in a subset of RB and NB cell lines and tumors (37–39). *MYCN* amplification is common in stage IV NB tumors and is a well documented indicator of poor prognosis. A general trend toward a poorer clinical prognosis is observed when both the *MYCN* and *DDX1* genes are amplified compared with when only *MYCN* is amplified (38, 39), suggesting a possible role for DDX1 in NB tumor formation or progression.

It is generally accepted that co-amplified genes are not overexpressed unless they provide a selective growth advantage to the cell (48, 49). For example, although *ERBA* is closely linked to *ERBB2* in breast cancer and both genes are commonly amplified in these tumors, *ERBA* is not overexpressed (48). Similarly, three genes mapping to 12q13–14 (*CDK4*, *SAS*, and *MDM2*) are overexpressed in a high percentage of malignant gliomas showing amplification of this chromosomal region, while other genes mapping to this region (*GADD153*, *GLI*, and *A2MR*) are rarely overexpressed in gene-amplified malignant gliomas (50, 51). The first three genes are probably the main targets of the amplification process, while the latter three genes are probably incidentally included in the amplicons. The data shown here indicate that DDX1 is overexpressed at both the protein and RNA levels in *DDX1*-amplified RB and NB cell lines and that there is a strong correlation between *DDX1* gene copy number, *DDX1* RNA levels, and DDX1 protein levels in these lines. Our results are therefore consistent with DDX1 overexpression playing a positive role in some aspect of NB and RB tumor formation or progression. Recently, Weiss *et al.* (52) have shown that transgenic mice that overexpress MYCN develop NB tumors several months after birth. They conclude that MYCN overexpression can contribute to the initiation of tumorigenesis but that additional events are required for tumor formation. Amplification of *DDX1* may represent one of many alternative pathways by which a normal precursor "neuroblast" or "retinoblast" cell gains malignant properties.

The function of the majority of tissue-specific or developmentally regulated DEAD box genes remains unknown. However, some members of this protein family have been either directly or indirectly implicated in tumorigenesis. For example, the p68 gene has been found to be mutated in the ultraviolet light-induced murine tumor 8101 (53), while DDX6 (also known as RCK or p54) is encoded by a gene located at the breakpoint of the translocation involving chromosomes 11 and 14 in a cell line derived from a B-cell lymphoma (54, 55). Similarly, the production of a chimeric protein between *DDX10* and the nucleoporin gene *NUP98* has been proposed to be involved in the pathogenesis of a subset of myeloid malignancies with inv(11) (p15q22) (56). Interestingly, Grandori *et al.* (57) have shown that MYCC interacts with a DEAD box gene called *MrDb*, suggesting that the transcription of some DEAD box genes could be regulated through interaction with members of the MYC family. Future work will involve determining whether DDX1 represents another member of the DEAD box family with a role in the tumorigenic process.

DEAD box proteins have been implicated in translation initiation, RNA splicing, RNA degradation, and RNA stability (3, 18, 19). We carried out subcellular localization studies in an attempt to obtain a general indication of the function of DDX1. We found DDX1 protein in both the cytoplasm and nucleus of *DDX1*-amplified NB and RB lines. In contrast, DDX1 was

mainly located in the nucleus of nonamplified cell lines and normal fibroblast cultures. DDX1 was not associated with cellular organelles or with membranes based on immunoelectron microscopy. We therefore propose that the primary role of DDX1 is in the nucleus. The presence of DDX1 in the cytoplasm of *DDX1*-amplified cells may indicate that the amount of DDX1 protein that is allowed in the nucleus is tightly regulated. Alternatively, DDX1 may play a dual role in the nucleus and cytoplasm of *DDX1*-amplified cells.

An important component of our analysis was to identify the translation and transcription initiation sites of *DDX1*. We used a combination of techniques to identify the transcription start site: screening of RB and fetal brain libraries, RACE, primer extension, genomic DNA sequencing, S1 nuclease mapping, and Northern blot analysis using probes to the predicted 5′-end of the transcript. The transcription start site identified using these techniques is located ~300 nt upstream of the predicted translation initiation codon and was readily detected in three *DDX1*-amplified lines and barely detectable in a fourth amplified line. The 5′-untranslated region as well as the first in frame methionine are encoded within the first exon of *DDX1*. An in frame stop codon is located 123 nt upstream of the predicted initiation codon. We were unable to identify the transcription initiation site of *DDX1* in two of the six amplified lines tested as well as in nonamplified lines. Although it remains possible that there are different transcription start sites in different cell lines, detection of lower levels (rather than the absence) of the 5′-most 160 nt of the *DDX1* transcript in IMR-32, Y79, and LA-N-5 compared with RB522A, BE(2)-C, and LA-N-1 supports a quantitative rather than a qualitative difference in the 5′-end of this transcript in these cells. Our results suggest that the 5′-end of *DDX1* mRNA is rarely intact, even in mRNA preparations that otherwise appear to be of high quality based on analysis of control transcripts. The 5′-end of *DDX1* mRNA may therefore be especially susceptible to degradation, perhaps because of its sequence and/or secondary structure.

In conclusion, we have mapped the 5′-end of the 2.7-kb *DDX1* transcript and have identified the predicted translation initiation site of DDX1 protein. We have found that *DDX1*-amplified RB and NB tumor lines overexpress DDX1 protein and that there is a good correlation between gene copy number and both transcript and protein levels in these cells. We have shown that DDX1 protein is primarily located in the nucleus of cells that are not *DDX1*-amplified. In contrast, DDX1 is present in both the nucleus and cytoplasm of *DDX1*-amplified NB and RB lines. A cytoplasmic location in *DDX1*-amplified lines may indicate that the amount of nuclear DDX1 is tightly regulated or that DDX1 plays a dual role in the cytoplasm and nucleus of these cells.

Acknowledgments—We thank Walter Dixon, Brenda Gallie, Ajay Pandita, Jeremy Squire, and Herman Yeger for the neuroblastoma and retinoblastoma cell lines. We thank Halyna Marusyk for carrying out the electron microscopy analyses. We are grateful to Randy Andison for expert help in the preparation of the manuscript and to Stacey Hume for helpful discussions.

REFERENCES

1. Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., and Slonimski, P. P. (1989) *Nature* **337,** 121–122
2. Wassarman, D. A., and Steitz, J. A. (1991) *Nature* **349,** 463–464
3. Schmid, S. R., and Linder, P. (1992) *Mol. Microbiol.* **6,** 283–292
4. Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C., and Sonenberg, N. (1990) *Mol. Cell. Biol.* **10,** 1134–1144
5. Hirling, H., Scheffner, M., Restle, T., and Stahl, H. (1989) *Nature* **339,** 562–564
6. Liang, L., Diehl-Jones, W., and Lasko, P. (1994) *Development* **120,** 1201–1211
7. Gururajan R., Mathews, L., Longo, F., and Weeks, D. L. (1994) *Proc. Natl. Acad. Sci.* U. S. A. **91,** 2056–2060
8. Dalbadie-McFarland, G., and Abelson, J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4236–4240
9. Jamieson, D. J., Rahe, B., Pringle, J., and Beggs, J. D. (1991) *Nature* **349,** 715–717
10. Strauss, E. J., and Guthrie, C. (1994) *Nucleic Acids Res.* **22,** 3187–3193
11. Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T., and Schnier, J. (1988) *Nature* **336,** 496–498
12. Leroy, P., Alzari, P., Sassoon, D. Wolgemuth, D., and Fellous, M. (1989) *Cell* **57,** 549–559
13. Hay, B., Jan, L. Y., and Jan, Y. N. (1988) *Cell* **55,** 577–587
14. De Valoir, T., Tucker, M., Belikoff, E. J., Camp, L. A. Bolduc, C., and Beckingham, K. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 2113–2117
15. Ford, M. J., Anton, I. A., and Lane, D. P. (1988) *Nature* **332,** 736–738
16. Iggo, R. D., and Lane, D. P. (1989) *EMBO J.* **8,** 1827–1831
17. Buelt, M. K., Glidden, B. J., and Storm, D. R. (1994) *J. Biol. Chem.* **269,** 29367–29370
18. Iost, I., and Dreyfus, M. (1994) *Nature* **372,** 193–196
19. Py, B., Higgins, C. F., Krisch, H. M., and Carpousis, A. J. (1996) *Nature* **381,** 169–172
20. Missel, A., Souza, A. E., Nörskau, G., and Göringer, H. U. (1997) *Mol. Cell. Biol.* **17,** 4895–4903
21. Godbout, R., and Squire, J. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 7578–7582
22. Kiledjian, M., and Dreyfuss, G. (1992) *EMBO J.* **11,** 2655–2664
23. Godbout, R., Hale, M., and Bisgrove, D. (1994) *Gene (Amst.)* **138,** 243–245
24. Blackwood, E. M., and Eisenman, R. N. (1991) *Science* **251,** 1211–1217
25. Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I., and Land, H. (1992) *Nature* **359,** 423–426
26. Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E., and Bishop, J. M. (1984) Science **224,** 1121–1124
27. Seeger, R. C., Brodeur, G. M., Sather, H., Dalton, A., Siegel, S. E., Wong, K. Y., and Hammond, D. (1985) *N. Engl. J. Med.* **313,** 1111–1116
28. Cohn, S. L., Salwen, H., Quasney, M. W., Ikegaki, N., Cowan, J. M., Herst, C. V., Sharon, B., Kennett, R. H., and Rosen, S. T. (1991) *Prog. Clin. Biol. Res.* **366,** 21–27
29. Cowell, J. K. (1982) *Annu. Rev. Genet.* **16,** 21–59
30. Zehnbauer, B. A., Small, D., Brodeur, G. M. Seeger, R., and Vogelstein, B. (1988) *Mol. Cell. Biol.* **8,** 522–530
31. Akiyama, K., and Nishi, Y. (1991) *Nucleic Acids Res.* **19,** 6887–6894
32. Schneider, S. S., Hiemstra, J. L., Zehnbauer, B. A., Taillon-Miller, P., Le Paslier, D., Vogelstein, B., and Brodeur, G. M. (1992) *Mol. Cell. Biol.* **12,** 5563–5570
33. Amler, L. C., Schürmann, J., and Schwab, M. (1996) *Genes Chromosomes Cancer* **15,** 134–137
34. Kuroda, H., White, P. S., Sulman, E. P., Manohar, C., Reiter, J. L., Cohn, S. L., and Brodeur, G. M. (1996) *Oncogene* **13,** 1561–1565
35. Noguchi, Y., Akiyama, K., Yokoyama, M., Kanda, N., Matsunaga, T., and Nishi, Y. (1996) *Genes Chromosomes Cancer* **15,** 129–133
36. Pandita, A., Godbout, R., Zielenska, M., Thorner, P., Bayani, J., and Squire, J. A. (1997) *Genes Chromosomes Cancer* **20,** 243–252
37. Manohar, C. F., Salwen, H. R., Brodeur, G. M., and Cohn, S. L. (1995) *Genes Chromosomes Cancer* **14,** 196–203
38. Squire, J. A., Thorner, P. S., Weitzman, S., Maggi, J. D., Dirks, P., Doyle, J., Hale, M., and Godbout, R. (1995) *Oncogene* **10,** 1417–1422
39. George, R. E., Kenyon, R. M., McGuckin, A. G., Malcolm, A. J., Pearson, A. D. J., and Lunec, J. (1996) *Oncogene* **12,** 1583–1587
40. Reiter, J. L., and Brodeur, G. M. (1996) *Genomics* **32,** 97–103
41. Favaloro, J., Treisman, R., and Kamen, R. (1980) *Methods Enzymol.* **65,** 718–745
42. Dignam, J. D. (1990) *Methods Enzymol.* **182,** 194–203
43. Graham, J. (1984) in *Centrifugation: A Practical Approach* (Rickwood, D. ed.) 2nd Ed., pp. 161–182, IRL Press, Oxford
44. Godbout, R., Marusyk, H., Bisgrove, D., Dabbagh, L., and Poppema, S. (1995) *Exp. Eye Res.* **60,** 645–657
45. Miller, K. G., and Sollner-Webb, B. (1981) *Cell* **27,** 165–174
46. Kozak, M. (1988) *Nucleic Acids Res.* **15,** 8125–8132
47. Rafti, F., Scarvelis, D., and Lasko, P. F. (1996) *Gene (Amst.)* **171,** 225–229
48. Van de Vijver, M., van de Bersselaar, R., Devilee, P., Cornelisse, C., Peterse, J., and Nusse, R. (1987) *Mol. Cell. Biol.* **7,** 2019–2023
49. Gaudray, P., Szepetowski, P., Escot, C., Birnbaum, D., and Theillet, C. (1992) *Mutat. Res.* **276,** 317–328
50. Forus, A., Florence, V. A., Maelandsmo, G. M., Meltzer, P. S., Fodstad, O., and Myklebost, O. (1993) *Cell Growth Differ.* **4,** 1065–1070
51. Reifenberger, G., Reifenberger, J., Ichimura, K., Meltzer, P. S., and Collins, V. P. (1994) *Cancer Res.* **54,** 4299–4303
52. Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G., and Bishop, J. M. (1997) *EMBO J.* **16,** 2985–2995
53. Dubey, B. P., Hendrickson, R. C., Meredith, S. C., Siegel, C. T., Shabanowitz, J., Skipper, J. C. A., Engelhard, V. H., Hunt, D. F., and Schreiber, H. (1997) *J. Exp. Med.* **185,** 695–705
54. Akao, Y., Seto, M., Yamamoto, K., Iida, S., Nakazawa, S., Inazawa, J., Abe, T., Takahashi, T., and Ueda, R. (1992) *Cancer Res.* **52,** 6083–6087
55. Lu, D., and Yunis, J. J. (1992) *Nucleic Acids Res.* **20,** 1967–1972
56. Arai, Y., Hosoda, F., Kobayashi, H., Arai, K., Hayashi, Y., Kamada, N., Kaneko, Y., and Ohki, M. (1997) *Blood* **89,** 3936–3944
57. Grandori, C., Mac, J., Siëbelt, F., Ayer, D. E., and Eisenman, R. N. (1996) *EMBO J.* **15,** 4344–4357

[CANCER RESEARCH 61, 2409–2412, March 15, 2001]

Advances in Brief

BMI-1 Gene Amplification and Overexpression in Hematological Malignancies Occur Mainly in Mantle Cell Lymphomas[1]

Silvia Beà, Frederic Tort, Magda Pinyol, Xavier Puig, Luis Hernández, Silvia Hernández, Pedro L. Fernández, Maarten van Lohuizen, Dolors Colomer, and Elias Campo[2]

The Hematopathology Section, Laboratory of Anatomic Pathology, Hospital Clínic, Institut d'Investigacions Biomèdiques "August Pi i Sunyer" (IDIBAPS), University of Barcelona, 08036 Barcelona, Spain [S. B., F. T., M. P., X. P., L. H., S. H., P. L. F., D. C., E. C.], and Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands [M. v. L.]

Abstract

The *BMI-1* gene is a putative oncogene belonging to the Polycomb group family that cooperates with c-*myc* in the generation of mouse lymphomas and seems to participate in cell cycle regulation and senescence by acting as a transcriptional repressor of the *INK4a/ARF* locus. The *BMI-1* gene has been located on chromosome 10p13, a region involved in chromosomal translocations in infant leukemias, and amplified in occasional non-Hodgkin's lymphomas (NHLs) and solid tumors. To determine the possible alterations of this gene in human malignancies, we have examined 160 lymphoproliferative disorders, 13 myeloid leukemias, and 89 carcinomas by Southern blot analysis and detected *BMI-1* gene amplification (3- to 7-fold) in 4 of 36 (11%) mantle cell lymphomas (MCLs) with no alterations in the *INK4a/ARF* locus. BMI-1 and $p16^{INK4a}$ mRNA and protein expression were also studied by real-time quantitative reverse transcription-PCR and Western blot, respectively, in a subset of NHLs. BMI-1 expression was significantly higher in chronic lymphocytic leukemia and MCL than in follicular lymphoma and large B cell lymphoma. The four tumors with gene amplification showed significantly higher mRNA levels than other MCLs and NHLs with the *BMI-1* gene in germline configuration. Five additional MCLs also showed very high mRNA levels without gene amplification. A good correlation between BMI-1 mRNA levels and protein expression was observed in all types of lymphomas. No relationship was detected between BMI-1 and $p16^{INK4a}$ mRNA levels. These findings suggest that *BMI-1* gene alterations in human neoplasms are uncommon, but they may contribute to the pathogenesis in a subset of malignant lymphomas, particularly of mantle cell type.

Introduction

The *BMI-1*[3] gene is a putative oncogene of the Polycomb group originally identified by retroviral insertional mutagenesis in Eμ-c-myc transgenic mice infected with the Moloney murine leukemia virus (1, 2). These animals had a rapid development of pre-B cell lymphomas showing frequent proviral insertions near the *BMI-1* gene. This integration resulted in BMI-1 overexpression suggesting a cooperative effect between *C-MYC* and *BMI-1* genes in the development of these tumors (3, 4). Recent studies have indicated that the *BMI-1* gene may also participate in cell cycle control and senescence through the *INK4a/ARF* locus by acting as an upstream negative regulator of $p16^{INK4a}$ and $p14/p19^{ARF}$ gene expression (5). The human *BMI-1* gene has been mapped to chromosome 10p13 (6), a region involved in chromosomal translocations in infant leukemias (7) and rearrangements in malignant T cell lymphomas (8, 9). More recently, high-level DNA amplifications of this region have been found by comparative genomic hybridization in NHLs and solid tumors (10, 11). However, the possible implication of the *BMI-1* gene in these alterations and its role in the pathogenesis of human tumors is not known. The aim of this study was to analyze the possible *BMI-1* gene alterations and expression in a large series of human neoplasms and to determine the relationship with *INK4a/ARF* locus aberrations.

Materials and Methods

Case Selection. A series of 262 human tumors, including 173 hematological malignancies and 89 carcinomas (Table 1), matched normal tissues from all carcinomas, 11 samples of normal peripheral mononuclear cells, and 5 reactive lymph nodes and tonsils, were selected based on the availability of frozen samples for molecular analysis.

DNA Extraction and Southern Blot Analysis. Genomic DNA was obtained using Proteinase K/RNase treatment. 15 μg were digested with *Eco*RI and *Hin*dIII restriction enzymes (Life Technologies, Inc., Gaithersburg, MD), for Southern blot analysis and hybridized with a 1.5-kb *Pst*I fragment of the partial *BMI-1* cDNA (6).

RNA Extraction and Real-time Quantitative RT-PCR. Total RNA was obtained from 67 lymphoid neoplasms (10 CLLs, 27 MCLs, 8 FLs, and 22 LCLs) using guanidine/isothiocyanate extraction and cesium/chloride gradient centrifugation. One μg of total RNA was transcribed into cDNA using MMLV-reverse transcriptase (Life Technologies, Inc.) and random hexamers, following manufacturer's directions. Sequences of the *BMI-1* and the *p16* detection probes and primers were designed using the Primer Express program (Applied Biosystems, Foster City) as follows: *BMI-1* sense, 5′-CTGGTTGCCCATTGACAGC-3′; *BMI-1* antisense, 5′-CAGAAAATGAATGCGAGCCA-3′; *p16* sense, 5′-CAACGCACCGAATAGTTACGG-3′; *p16* antisense, 5′-AACTTCGTCCTCCAGAGTCGC-3′. The probes *BMI-1*, 5′-CAGCTCGCTTCAAGATGGCCGC-3′, and *p16*, 5′-CGGAGGCCGATCCAGGTGGGTA-3′, were labeled with 6-carboxy-fluorescein as the reporter dye. The TaqMan-GAPDH Control Reagents (Applied Biosystems) were used to amplify and detect the *GAPDH* gene, as recommended by the manufacturer. The quantitative assay amplified 1 μml of cDNA in two to four replicates using the primers and probes described above and the standard master mix (Applied Biosystems). All reactions were performed in an ABI PRISM 7700 Sequence Detector System (Applied Biosystems). GAPDH, BMI-1, and $p16^{INK4a}$ expression was related to a standard curve derived from serial dilutions of Raji cDNA. The RUs of BMI-1 and $p16^{INK4a}$ expression were defined as the mRNA levels of these genes normalized to the GADPH expression level in each case.

Protein Analysis. Whole-cell protein extracts were obtained from additional frozen tissue available in 31 cases (7 CLLs, 12 MCLs, 8 FLs, and 4 LCLs), loaded onto a 10% SDS-polyacrylamide gel, and electroblotted to a nitrocellulose membrane (Amersham). Blocked membranes were incubated sequentially with the monoclonal antibody BMI-F6 (12), antimouse conju-

Received 10/16/00; accepted 1/29/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported by Grant SAF 99/20 from Comision Interministerial de Ciencia y Tecnologia, European Union Contract QLG1-CT-2000-689, the Asociación Española contra el Cáncer, and Generalitat de Catalunya 98SGR21. S. B. and F. T. were fellows supported by Spanish Ministerio de Educación y Cultura, and S. H. was supported by the Asociación Española contra el Cáncer and the Fundació Rius i Virgili.

[2] To whom requests for reprints should be addressed, at the Department of Pathology, Hospital Clínic, University of Barcelona, Villarroel 170, 08036-Barcelona, Spain. Phone: 34 93 227 5450; Fax: 34 93 227 5572; E-mail: campo@medicina.ub.es.

[3] The abbreviations used are: BMI-1, B cell-specific Moloney murine leukemia virus integration site 1; NHL, non-Hodgkin's lymphoma; CLL, chronic lymphocytic leukemia; FL, follicular lymphoma; LCL, large B cell lymphoma; MCL, mantle cell lymphoma; RT-PCR, reverse-transcription-PCR; RU, relative units.

Table 1 *Hematological malignancies and solid tumor samples analyzed for BMI-1 gene alterations*

Tissue samples	No. of cases
Hematological malignancies	
Hodgkin's disease	2
B cell lymphoproliferative disorders	
B-Acute lymphoblastic leukemia	14
CLL	29
Hairy cell leukemia	4
FL	15
MCL	36
LCL	40
T cell lymphoproliferative disorders	
T-Acute lymphoblastic leukemia	8
Large granular cell leukemia	4
Peripheral T-cell lymphoma	8
Myeloproliferative disorders	
Acute myeloid leukemia	7
Chronic myeloid leukemia	6
Solid tumors	
Colon carcinoma	26
Breast carcinoma	29
Laryngeal squamous cell carcinoma	34
Total	262

gated to horseradish peroxidase (Amersham), and detected by enhanced chemiluminescence (Amersham) according to the manufacturer's recommendations.

Statistical Analysis. Because of the non-normal distribution of the samples and the small size of some subsets of tumors, the statistical evaluation was performed using nonparametric tests (SPSS, version 9.0). Comparison between mRNA expression levels in the different groups of NHLs was performed using the Kruskal-Wallis Test, with a P for significance set at 0.05. For differences between particular groups, the conservative Bonferroni procedure was performed, and the P was set at 0.005. The remaining statistical analyses were carried out using the Mann-Whitney nonparametric U test (significance, $P < 0.05$). The comparison between BMI-1 and $p16^{INK4a}$ quantitative mRNA levels was also performed using the Pearson's correlation coefficient.

Results

***BMI-1* Gene Amplification.** The *BMI-1* gene was examined by Southern blot in a large series of human tumors and normal samples (Table 1). The cDNA probe used in the study detected three *Eco*RI fragments of 7.3, 3.8, and 2.6 kb and three *Hin*dIII fragments of 6.2, 4, and 3.5 kb. *BMI-1* gene amplification (3- to 7-fold) was detected in 4 of 36 (11%) MCLs (Fig. 1). The amplifications were confirmed with both restriction enzymes. The amplified MCLs were two blastoid and two typical variants. No amplifications were observed in any of the solid tumors when compared with their respective matched non-neoplastic mucosa. No *BMI-1* gene rearrangements were observed in any of the samples examined.

BMI-1 mRNA Expression. To determine the BMI-1 expression pattern in NHL we analyzed BMI-1 mRNA levels by real-time quantitative RT-PCR in 67 lymphomas (10 CLLs, 27 MCLs, 8 FLs, and 22 LCLs), including the four tumors with gene amplification. A distinct BMI-1 mRNA expression pattern was observed in the different types of lymphomas (Fig. 2; Kruskal-Wallis Test; $P < 0.001$). The BMI mRNA levels in CLLs (mean, 2.2 RU; SD, 1.3) and MCLs with no *BMI-1* gene amplification (mean, 2.5 RU; SD, 2.3) were significantly higher than in FLs (mean, 0.9 RU; SD, 0.8) and LCLs (mean, 0.6 RU; SD, 0.4; Mann-Whitney nonparametric U test; $P < 0.01$). The 4 MCLs with *BMI-1* gene amplification showed significantly higher levels of expression than all other groups of tumors (mean, 5.1 RU; SD, 1.6; $P < 0.005$). In addition, five typical MCLs with no structural alterations of the gene also showed very high levels of BMI-1 mRNA expression ranging from 4 to 9.8 RU, similar to cases with gene amplification (Fig. 2*A*).

BMI-1 Protein Expression. BMI-1 protein expression was examined by Western blot in 31 tumors (7 CLLs; 12 MCLs, including two cases with *BMI-1* gene amplification and 4 cases with mRNA overexpression and no structural alteration of the gene; 8 FLs, and 4 LCLs) in which additional frozen tissue was available. The monoclonal antibody against BMI-1 detected three closely migrating proteins of M_r 45,000–48,000 (2). The two more slowly migrating bands probably represent phosphorylated isoforms of the protein (12). The two MCLs with gene amplification and three of four cases with mRNA overexpression without amplification of the gene showed very high levels of protein expression. The remaining MCLs and CLLs showed intermediate levels of protein expression, whereas low- or no-expression signals were detected in the LCLs and FLs included in the study (Fig. 3). These results indicate that BMI-1 protein expression in NHL is concordant with the mRNA levels observed by real-time quantitative RT-PCR.

Relationship between BMI-1 and $p16^{INK4a}$ Gene Alterations. The *INK4a/ARF* locus has been recently identified as a downstream target of the transcriptional repressing activity of the *BMI-1* gene, suggesting that this gene may contribute to human neoplasias with wild type *INK4/ARF* (5). Most of the lymphoproliferative disorders analyzed in the present study, including the four cases with *BMI-1* gene amplification, had been previously examined for *p53* gene mutations and *INK4a/ARF* locus alterations, including gene deletions, mutations, hypermethylation, and expression (13, 14). The four MCLs with *BMI-1* gene amplification and mRNA overexpression and the five tumors with BMI-1 mRNA overexpression with no structural alterations of the gene showed a wild-type configuration of the *INK4a/ARF* locus (13). However, one case with *BMI-1* gene amplification and one case with mRNA overexpression with no alteration of the gene showed *p53* gene mutations associated with allelic deletions.

To determine the possible relationship between BMI-1 and $p16^{INK4a}$ mRNA expression, $p16^{INK4a}$ mRNA levels were evaluated by real-time quantitative RT-PCR in 50 tumors (10 CLLs, 27 MCLs, and 13 LCLs), including 6 cases with alterations in the *INK4a/ARF* locus (2 MCLs and 1 LCL with *p16*^{INK4a} gene deletion, 2 LCLs with *p16* promoter hypermethylation, and 1 CLL with *p16*^{INK4a} gene mutation), and the 4 lymphomas with *BMI-1* amplification. Negative or negligible levels of $p16^{INK4a}$ were observed in the 6 tumors with *INK4a/ARF* locus alterations. These cases were not included in the comparisons between BMI-1 and $p16^{INK4a}$ mRNA expression. The $p16^{INK4a}$ expression levels were relatively similar in the different types of tumors. Only LCLs tended to have lower levels of expression, but the differences did not reach statistical significance (Fig. 2*B*). No differences were observed in the $p16^{INK4a}$ mRNA levels between tumors with *BMI-1* gene amplification and overexpression and lymphomas with germline configuration of the gene.

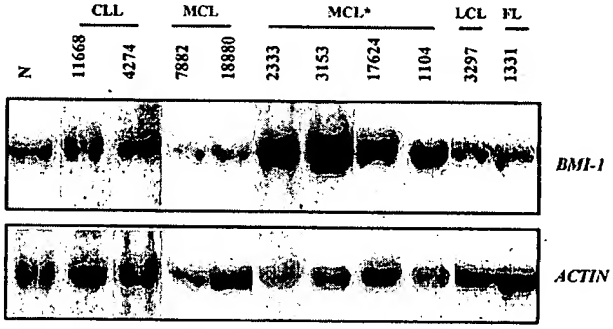

Fig. 1. Southern blot analysis of *BMI-1* gene. Four MCLs (*MCL**) showed *BMI-1* gene amplification (3- to 7-fold) compared with non-neoplastic tissues (*N*) and other NHLs. No amplifications or gene rearrangements were detected in the remaining NHLs and carcinomas included in the study.

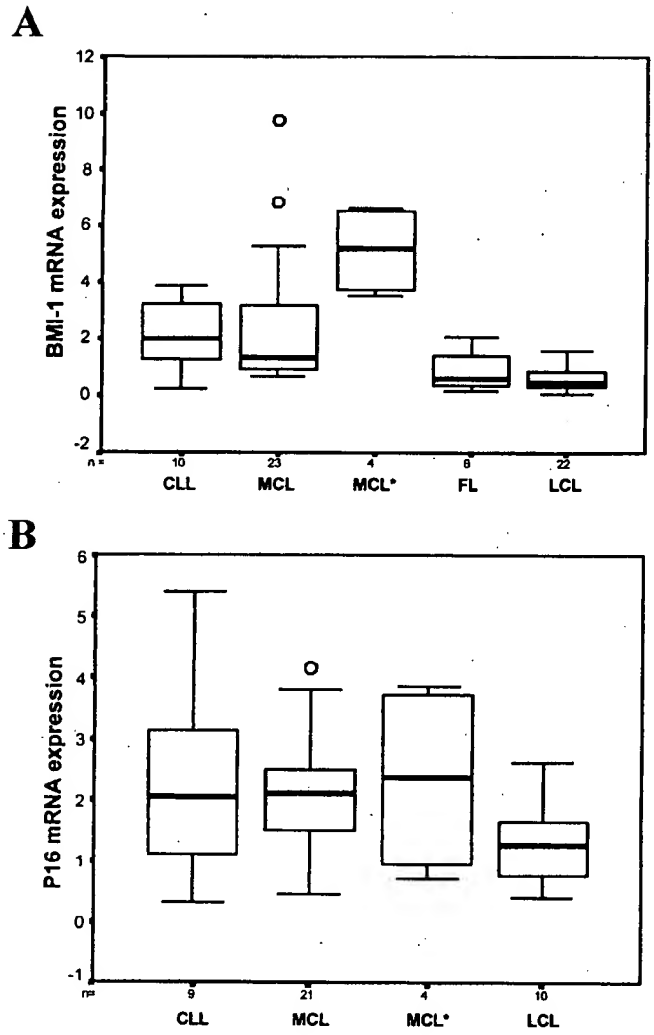

Fig. 2. *A*, quantitative BMI-1 mRNA transcript analysis (median and range) using real-time RT-PCR in a series of NHLs. MCLs with *BMI-1* gene amplification (*MCL**) revealed significantly higher overall BMI-1 mRNA levels than all other types of NHLs, including MCLs with no structural alterations of the gene ($P < 0.005$). MCLs and CLLs expressed significantly higher levels than FLs and LCLs ($P < 0.001$). Results are depicted as the ratio of absolute BMI-1:GADPH mRNA transcript numbers (RU). *Bars*, SD. *B*, quantitative $p16^{INK4a}$ mRNA transcript analysis (median and range) using real-time RT-PCR in a series of NHLs. Expression levels were relatively similar in the different types of tumors. Results are depicted as the ratio of absolute $p16^{INK4a}$:GADPH mRNA transcript numbers (RU). *Bars*, SD.

Discussion

In the present study, we have examined a large series of human tumors for the presence of gene alterations and mRNA expression of the *BMI-1* gene. Gene amplification was identified in four MCLs. These tumors showed significantly higher levels of mRNA and protein expression compared with other lymphomas with *BMI-1* in germline configuration. BMI-1 expression levels were also highly upregulated in a subset of MCLs with no apparent structural alterations of the gene. No alterations were detected in any of the different types of carcinomas included in the study. *BMI-1* is considered an oncogene belonging to the Polycomb group family of genes. These proteins mainly act as transcriptional regulators, controlling specific target genes involved in development, cell differentiation, proliferation, and senescence. Different studies have shown the implication of BMI-1 overexpression in the development of lymphomas in murine and feline animal models (3, 4). The findings of the present study indicate for the first time that *BMI-1* gene alterations in human neoplasms are an uncommon phenomenon, but they seem to occur mainly in a subset of NHLs, particularly of mantle cell type.

The human *BMI-1* gene has been mapped to chromosome 10p13. High-level DNA amplifications and gains in this region have been identified by comparative genomic hybridization in occasional solid tumors and NHLs (10, 11). Different chromosomal translocations involving the 10p13 region have also been identified in infant leukemias and T cell lymphoproliferative disorders (7, 8, 15). Most acute leukemias with this chromosomal alteration occur in children <12 months of age, whereas it seems to be extremely rare in adults. 10p translocations in T-cell lymphoproliferative disorders have been observed mainly in adult T cell leukemia/lymphomas and occasional cutaneous T cell lymphomas. In our study, we did not observe *BMI-1* rearrangements or amplifications in any of the acute leukemias or T cell lymphomas. However, all of the acute leukemias in this study were diagnosed in patients over 16 years, and no adult T cell leukemia/lymphomas or cutaneous lymphomas could be included in the series. Similarly, high-level DNA amplifications at the 10p13 region have been detected in head and neck carcinomas and other solid tumors. Although we found no evidence for *BMI-1* gene rearrangements or amplifications in a substantial set of carcinomas, this does not exclude the possibility of increased gene expression or protein levels in these tumors. Additional studies are required to elucidate the possible involvement of *BMI-1* in these particular groups of human neoplasms.

In human hematopoietic cells, BMI-1 is preferentially expressed in primitive CD34+ bone marrow cells, whereas it is negative or very low in more mature CD34− cells (16). In peripheral lymphocytes, and particularly in follicular B cells, BMI-1 protein expression has been detected in resting cells of the mantle zone, whereas it is downregulated in proliferating germinal center cells (17, 18). These observations indicate that BMI-1 expression in normal hematopoietic cells is tightly regulated in relation with cell differentiation in bone marrow and antigen-specific response in peripheral lymphocytes. BMI-1 expression in human tumors has not been examined previously. In this study, we have demonstrated that BMI-1 mRNA and protein expression show a distinct pattern in different types of lymphomas. Thus, BMI-1 levels were low in LCLs and FLs and significantly higher in MCLs and CLLs. These findings suggest that BMI-1 expression patterns in B cell lymphomas maintain in part the expression profile of their normal cell counterparts; because FLs and at least a subgroup of LCLs are considered lymphomas derived from follicular germinal center cells, whereas MCLs and CLLs are tumors mainly derived from naive pregerminal center cells. However, the four MCLs with *BMI-1* gene amplification expressed significantly higher mRNA levels than all other tumors. In addition, five MCLs with no structural alterations of the gene showed high mRNA levels similar to those observed in tumors with *BMI-1* gene amplification, suggesting that other mechanisms may be involved in up-regulation of the gene in these lymphomas. Different studies using animal models have shown a dose-dependent effect of *BMI-1* gene expression on skeleton development

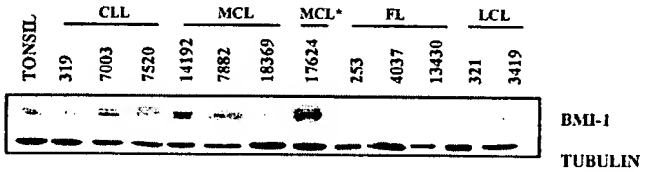

Fig. 3. Western blot analysis of BMI-1 protein in NHLs. The amplified MCL (17624) showed the highest BMI-1 protein levels, whereas other MCLs and CLLs had intermediate levels of expression. Very low or negative signal was observed in FLs and LCLs.

and lymphomagenesis (1, 3). These observations suggest that the high mRNA and protein levels detected in a subset of MCLs may play a role in the pathogenesis of these neoplasms.

Recent studies have identified the *INK4/ARF* locus as a downstream target of the BMI-1 transcriptional repressor activity, suggesting that BMI-1 overexpression may contribute to human neoplasias that retain the wild-type *INK4a/ARF* locus (5). Interestingly, in our study, *BMI-1* amplification and overexpression appeared in tumors with no alterations in $p16^{INK4a}$ and $p14^{ARF}$ genes. However, we could not detect differences in the expression levels of $p16^{INK4a}$ in tumors with and without *BMI-1* gene alterations. The reasons for this apparent discrepancy with experimental observations are not clear. One possibility may be that genes other than *INK4a/ARF* are the main targets of BMI-1 repressor activity in these tumors. Particularly, different genes of the HOX family are regulated by BMI-1 and may also be involved in lymphomagenesis (19, 20).

In conclusion, the findings of this study indicate that *BMI-1* gene expression is differentially regulated in B cell lymphomas. Alterations of the gene seem to be an uncommon phenomenon in human neoplasms, but they may contribute to the pathogenesis in a subset of MCLs. Although, *BMI-1* gene alterations occurred in tumors with wild-type *INK4a/ARF* locus, the possible cooperation between these genes and the oncogenic mechanisms of BMI-1 in human neoplasms require additional analysis.

Acknowledgments

The authors thank Iracema Nayach for her excellent technical assistance.

References

1. Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P., and Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E μ-myc transgenic mice. Cell, *65:* 753–763, 1991.
2. van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, A. Identification of cooperating oncogenes in E μ-myc transgenic mice by provirus tagging. Cell, *65:* 737–752, 1991.
3. Alkema, M. J., Jacobs, H., van Lohuizen, M., and Berns, A. Perturbation of B and T cell development and predisposition to lymphomagenesis in EμBmi1 transgenic mice require the Bmi1 RING finger. Oncogene, *15:* 899–910, 1997.
4. Haupt, Y., Bath, M. L., Harris, A. W., and Adams, J. M. *Bmi-1* transgene induces lymphomas and collaborates with myc in tumorigenesis. Oncogene, *8:* 3161–3164, 1993.
5. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A., and van Lohuizen, M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. Nature (Lond.), *397:* 164–168, 1999.
6. Alkema, M. J., Wiegant, J., Raap, A. K., Berns, A., and van Lohuizen, M. Characterization and chromosomal localization of the human proto-oncogene *BMI-1*. Hum. Mol. Genet., *2:* 1597–1603, 1993.
7. Pui, C. H., Raimondi, S. C., Murphy, S. B., Ribeiro, R. C., Kalwinsky, D. K., Dahl, G. V., Crist, W. M., and Williams, D. L. An analysis of leukemic cell chromosomal features in infants. Blood, *69:* 1289–1293, 1987.
8. Berger, R., Baranger, L., Bernheim, A., Valensi, F., Flandrin, G., and Berheimm, A. T. Cytogenetics of T-cell malignant lymphoma. Report of 17 cases and review of the chromosomal breakpoints. Cancer Genet. Cytogenet., *36:* 123–130, 1988.
9. D'Alessandro, E., Paterlini, P., Lo Re, M. L., Di Cola, M., Ligas, C., Quaglino, D., and Del Porto, G. Cytogenetic follow-up in a case of Sezary syndrome. Cancer Genet. Cytogenet., *45:* 231–236, 1990.
10. Bea, S., Ribas, M., Hernandez, J. M., Bosch, F., Pinyol, M., Hernandez, L., Garcia, J. L., Flores, T., Gonzalez, M., Lopez-Guillermo, A., Piris, M. A., Cardesa, A., Montserrat, E., Miro, R., and Campo, E. Increased number of chromosomal imbalances and high-level DNA amplifications in mantle cell lymphoma are associated with blastoid variants. Blood, *93:* 4365–4374, 1999.
11. Kriuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., el-Rifai, W., Hemmer, S., Wasenius, V. M., Vidgren, V., and Zhu, Y. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am. J. Pathol., *152:* 1107–1123, 1998.
12. Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van't Veer, L. J., Berns, A., and van Lohuizen, M. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. Genes Dev., *11:* 226–240, 1997.
13. Pinyol, M., Hernández, L., Martínez, A., Cobo, F., Hernández, S., Beà, S., López-Guillermo, A., Nayach, I., Palacín, A., Nadal, A., Fernández, P., Montserrat, E., Cardesa, A., and Campo, E. *INK4a/ARF* locus alterations in human non-Hodgkin's lymphomas mainly occur in tumors with wild type *p53* gene. Am. J. Pathol., *156:* 1987–1996, 2000.
14. Pinyol, M., Cobo, F., Beà, S., Jares, P., Nayach, I., Fernández, P. L., Montserrat, E., Cardesa, A., and Campo, E. *p16INK4a* gene inactivation by deletions, mutations, and hypermethylation is associated with transformed and aggressive variants of non-Hodgkin's lymphomas. Blood, *91:* 2977–2984, 1998.
15. Foot, A. B., Oakhill, A., and Kitchen, C. Acute monoblastic leukemia of infancy in Klinefelter's syndrome. Cancer Genet. Cytogenet., *61:* 99–100, 1992.
16. Lessard, J., Baban, S., and Sauvageau, G. Stage-specific expression of polycomb group genes in human bone marrow cells. Blood, *91:* 1216–1224, 1998.
17. Raaphorst, F. M., van Kemenade, F. J., Fieret, E., Hamer, K. M., Satijn, D. P., Otte, A. P., and Meijer, C. J. Cutting edge: polycomb gene expression patterns reflect distinct B cell differentiation stages in human germinal centers. J. Immunol., *164:* 1–4, 2000.
18. Raaphorst, F. M., van Kemenade, F. J., Blokzijl, T., Fieret, E., Hamer, K. M., Satijn, D. P., Otte, A. P., and Meijer, C. J. Coexpression of *BMI-1* and *EZH2* polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. Am. J. Pathol., *157:* 709–715, 2000.
19. Gould, A. Functions of mammalian Polycomb group and trithorax group related genes. Curr. Opin. Genet. Dev., *7:* 488–494, 1997.
20. van Oostveen, J., Bijl, J., Raaphorst, F., Walboomers, J., and Meijer, C. The role of homeobox genes in normal hematopoiesis and hematological malignancies. Leukemia, *13:* 1675–1690, 1999.

13

Gene Expression Analysis Using Microarrays

Sophie E. Wildsmith and Fiona J. Spence

13.1 Introduction

Microarrays are of increasing interest to both industry and academia as tools for 'gene hunting' and also as quantitative methods for routine analysis of large numbers of genes. Techniques such as real-time polymerase chain reaction (RT-PCR) TaqMan™ and SybrMan™ are generally considered to be more accurate, robust, larger in dynamic range and less capital intensive, but for rapid, large-scale gene expression analysis using limited mRNA, microarrays and gene chips are preferred.

13.2 Microarray experiments

Platforms

Global gene expression platforms are now available in multiple formats, including cDNA arrays, oligonucleotides spotted onto slides or *in situ* synthesised oligonucleotide arrays manufactured using photolithography. Commercial sources for these include Stratagene (La Jolla, CA), Memorec (Köln, Germany) and BD BioSciences (Oxford, UK) for cDNA microarrays, Mergen Ltd (San Leandro, CA) for spotted oligomers and Affymetrix (Palo Alto, CA) for oligoarrays synthesised *in situ*. Purchasing from a supplier is more expensive than generating microarrays in-house, although the latter is beneficial in labour-intensive institutions or when proprietary gene information is utilised.

Molecular Biology in Cellular Pathology. Edited by John Crocker and Paul G. Murray
 ISBN: 0-470-84475-2

'Off-the-shelf' microarrays may also have the advantage of rigorous quality control and standardised protocols.

It is possible to produce oligonucleotide spotted arrays in-house, by designing oligonucleotide sequences that match genes of interest and then purchasing purified oligonucleotides to spot down on glass or other substrates. Alternatively there are new systems such as that available from CombiMatrix (Mukilteo, WA, USA) for computer-aided design and *in situ* synthesis of oligonucleotides. However, production of cDNA microarrays is currently the most affordable and popular method and is now well established. Numerous sources of information on cDNA microarray fabrication are available in the literature and on the internet (Bowtell, 1999; Cheung *et al.*, 1999; Wildsmith and Elcock, 2001 and `http://cmgm.stanford.edu/pbrown`). Thus, this chapter will focus on the implementation of experiments and analysis of data from cDNA microarrays. The experimental procedure differs slightly according to the number of fluorophores (or channels) and the type and manufacturer of the array. We have attempted to describe a generic process, indicating where possible the different options. Figure 13.1 demonstrates the procedure for a two-colour hybridisation.

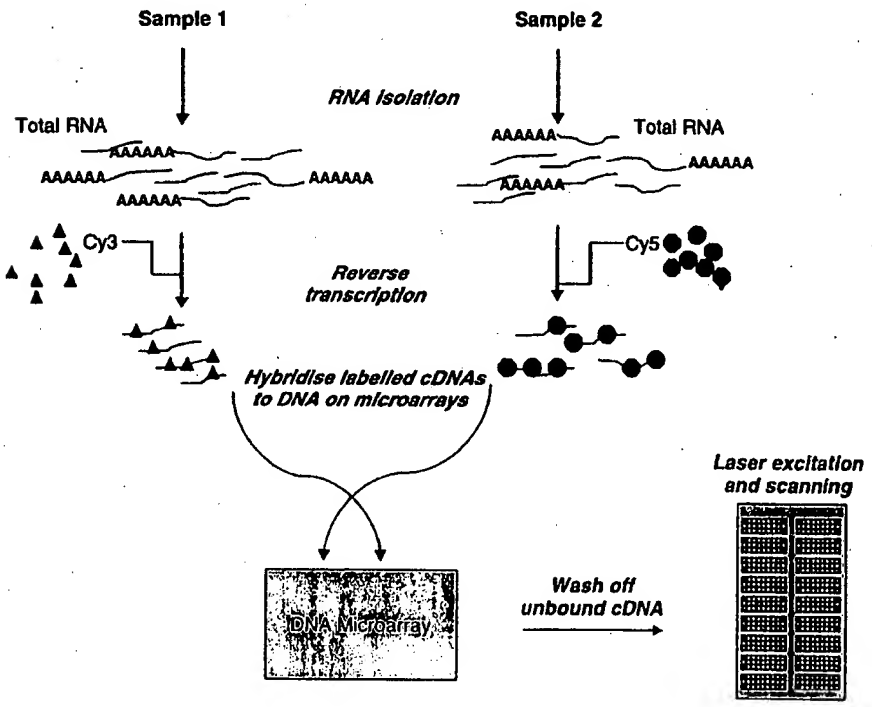

Figure 13.1 The microarray experimental process for two-colour hybridisations

RNA Extraction

First, RNA is extracted from the tissue or cells of interest. The quality of the RNA extracted is paramount to the overall success of the microarray experiment, as impurities in the sample can effect both the probe labelling efficiency and also stability of the fluorescent label (Hegde *et al.*, 2000). Snap-freezing of tissue in liquid nitrogen, immediately after harvesting, is used to preserve RNA integrity. Any further sectioning of the tissues should be carried out under RNAse-free conditions (Fernandez *et al.*, 1997). Total RNA can be extracted using kits such as TRIzol® (Invitrogen, Paisley, Scotland) and Rneasy (Qiagen, GmbH, Hilden, Germany). Some researchers perform a further extraction of mRNA; this results in a purer starting material but has the disadvantage of lower yields. Affymetrix recommend between 5 and 40 μg of total RNA is required for their GeneChips™ and 10 μg or less is the required amount of starting material for cDNA microarrays (Hegde *et al.*, 2000).

Sample Labelling

The mRNA is transcribed *in vitro*, with the concomitant inclusion of labelled nucleotides. The labels may be fluorescent or radioactive. In the case of dual channel/colour hybridisations, two samples will be labelled with dyes that fluoresce at different wavelengths, with different emission spectra. Example fluorophores, available coupled to nucleotides, are Cy3, Cy5, fluorescein and lissamine. Wildsmith *et al.* (2001) have demonstrated that AlexaFluor 546dUTP™ (Molecular Probes, Leiden, The Netherlands) gives a significantly higher signal than Cy3dCTP (Amersham Biosciences, Piscataway, NJ, USA). When performing two-colour hybridisations the control sample and 'test' sample are labelled with different fluorophores and the subsequent cDNA is then mixed together and hybridised simultaneously (Nuwaysir *et al.*, 1999). An advantage of simultaneously hybridising control and treated sample is that it obviates the need to control for differences in hybridisation conditions or between microarrays. A specific example of the huge impact this technique has had includes its use in the first published account of gene expression data of the entire genome of *Saccharomyces cerevisiae* (DeRisi *et al.*, 1997). In two-colour hybridisations, one would assume that the properties of the two fluorescent dyes being used are equivocal. In fact, for Cy5 and Cy3 this is not the case as Cy5 has been reported to give higher background fluorescence and also is more sensitive to photobleaching than Cy3 (Van Hal *et al.*, 2000). In addition, there is evidence from several independent sources that the combination of Cy3 and Cy5 dye labelling can affect data in certain genes. That is to say, when experiments are repeated and the dye combination for the two probes reversed, inconsistent results are obtained with certain genes (Taniguchi *et al.*, 2001). Despite these

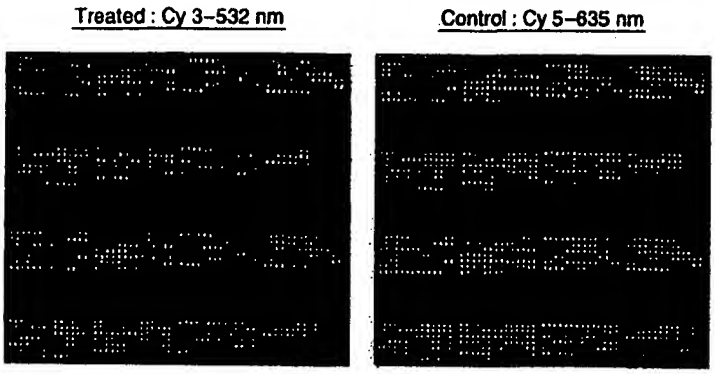

Figure 13.2 Two-colour fluorescent scan of human gene cDNA array. The probe mix consists of DNA from HepG2 control cells and cells treated with buthionine sulfoximine for 6 h. A colour version of this figure appears in the colour plate section

facts, two-colour hybridisations are widely accepted throughout the microarray community and an example of an image is shown in Figure 13.2.

Hybridisation and Processing

After labelling, the cDNA is purified (to remove unincorporated nucleotides), mixed with a hybridisation buffer and then applied to a cDNA microarray slide. The sample and the slide are heated prior to hybridisation in order to separate double-stranded DNA. A coverslip is applied (Shalon *et al.*, 1996), or preferably a hybridisation chamber is used to avoid evaporation and enable an even hybridisation. The hybridisation and subsequent wash steps are carried out at a buffer stringency and temperature that enables hybridisation of complementary strands of DNA but reduces non-specific binding.

Image Capture and Image Analysis

After hybridisation the microarray slides are scanned, using either a laser or a phosphorimager (depending on the type of label used). There are many different suppliers and models of fluorescence scanner, for example the ScanArray 5000 (Perkin Elmer Life Sciences, Zaventem, Belgium), GenePix 4000B (Axon GRI, Essex, UK) and the GeneArray® (Affymetrix, Santa Clara, CA). The choice of scanner is determined by sensitivity, resolution, flexible wavelength, file size generated, throughput and technical support available.

Images are analysed using software that measures the intensity of the signal from the hybridised spotted genes (spots), which provides a measurement of the amount of cDNA bound. Thus the initial concentration of messenger RNA is inferred. Early software packages 'drew' grids around the spots and

integrated across the whole area of the grid. This overcame problems associated with accurate location of the spots, which is problematic, especially if the spots on the printed arrays are poorly aligned. More recent versions of software 'draw' circles around the spots themselves and perform measurements within and outside of this boundary. For example, the background may be calculated from a region outside the spot boundary. The intensity of the signal from the spot may be calculated using median, mode or mean values of the pixels within the spot. Researchers differ in their preferences regarding using median or mean values (Hegde *et al.*, 2000) and this is likely to depend upon the protocols and software used.

Image analysis software commonly is supplied with scanners or can be bought from the same supplier. This has the advantage of being optimised for that specific type of microarray and the benefit of upgrades and technical support. Software for microarray analysis is available from BioDiscovery (`http://www.biodiscovery.com`), Imaging Research (`http://imagingresearch.com`), GenePix Pro (Amersham Biosciences, Piscataway, NJ), arraySCOUT 2.0 (`http://www.lionbioscience.com`), NIH (`http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html`), Stanford University (`http://rana.Stanford.EDU/software`) Media Cybernetics (Silver Spring, MD, USA) and TIGR (`http//www.tigr.org/softlab`). Important criteria for image analysis software include speed, ease of use, automation and the ability to distinguish artefact from real signal (Wildsmith and Elcock, 2001).

As the technology has evolved and more experience gained, it has become more and more apparent that the most significant issues facing microarray users are the processing of the vast quantities of data generated and deciding exactly what tools are the most appropriate for data analysis. Because of the enormity of this, we have dedicated a complete section to describing the current status of this area.

13.3 Data analysis

It is important to be cognisant of the fact that the practical laboratory aspects of using microarrays are only part of gene expression analysis. Many researchers generate vast volumes of data, without a clear understanding of how to manage and interpret them. Furthermore, the variability in microarray data confers additional problems for analysis. In some cases the purpose of the experiment will be a gene-hunting exercise, in which case a cursory indication of potential gene biomarkers is sufficient analysis. In other instances, such as pathway mapping and screening studies, it is paramount that results are statistically meaningful and valid. The next few sections detail some relatively simple analysis methods and recommendations for the benefit of researchers with minimal statistical

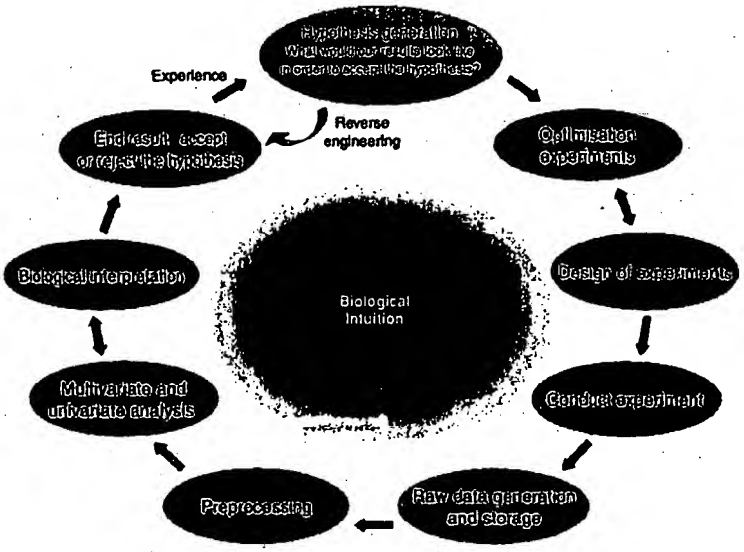

Figure 13.3 The ideal microarray experimental design and process

training. There are also suggestions for more advanced analysis for those who have the assistance of a statistician or specialist data analyst.

Most of the steps before, during and after performing a microarray experiment are optimally conducted with regard for statistics and data analysis. Careful planning before implementation facilitates the downstream analysis and interpretation of data. The following model summarises the entire microarray process with integration of the biological and data analysis components (Figure 13.3).

Hypothesis Generation

Any study is conceived for the purpose of investigating or obtaining supporting evidence for a biological hypothesis. Giving time at this early stage to consider downstream implications will pay dividends later. It is helpful if, rather than simply stating the aims of the experiment, the researcher asks the question 'What results do I expect?' or 'what answer will validate/invalidate my hypothesis?'. This 'reverse-engineering' proves useful in focusing the project, assessing the feasibility of the work, providing early preparation for data management and analysis and, importantly, in managing expectations with regard to outcomes.

A good example of careful experimental planning is demonstrated by Golub *et al.* (1999) in the classification of acute leukaemias in order to distinguish between acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML). Distinguishing between ALL and AML using conventional techniques is known to be a difficult task. The researchers maximised their probability of success by choosing an easier, more defined model (normal kidney vs. renal

cell carcinoma), on which to validate their analytical methods. In doing so they established that their techniques were suitable for classifying tissues according to disease and gained confidence in their approach before using the samples of real interest.

Optimisation Experiments

Although microarrays are becoming increasingly accessible to all, using these tools requires experience and it is unlikely that successful experiments will be conducted immediately. It is usual that some time is given to optimising a system for any specific application, for example for a given tissue or cell type. Additionally, the requirements for a given system may warrant some modifications. The standard approach for a scientist to take is to vary one parameter, whilst keeping all others constant. This is time-consuming and does not take into account the interactions between different factors. Well-designed, multifactorial experiments (Box *et al.*, 1978), provide a faster route for optimisation, with a statistical measure of confidence. An example of this technique is in the optimisation of microarray experimental conditions for preparation of fluorescent probes from rat liver tissue (Wildsmith *et al.*, 2001). When a major source of variation is revealed this can be investigated further with a view to minimising it or providing sufficient replicates to account for it.

Design of Experiments

Once confidence in the experimental procedure has been obtained the researcher is likely to have gained an insight into the reproducibility of the system. This assists in the design of the experiments, in particular in determining the minimal number of replicates necessary. Replication can be implemented at many stages – from biological samples through to microarray slides.

Owing to the enzyme-catalysed transcription reactions, a large amount of variation occurs during the probe-making stages in microarray experiments. Our work indicated that replicates should be made at this step and a minimum of six replicate probes are made for microarray experiments (Wildsmith *et al.*, 2001). These can be pooled or hybridised separately onto six microarray slides.

Lee *et al.* (2000) have examined the effect of the different location of cDNA spots on the glass slides and concluded that replicates are essential to provide meaningful data and to enable reliable inferences to be drawn.

With regard to commercially available gene chip systems, such as that available from Affymetrix (see Figure 13.4), the variation between chips, within a batch, is likely to be low due to stringent quality control and highly automated manufacture. The use of an automated wash station also reduces variability in intensities between chips. However, using a multi-step approach in the probe preparation and subsequent antibody binding steps may lead to variation between

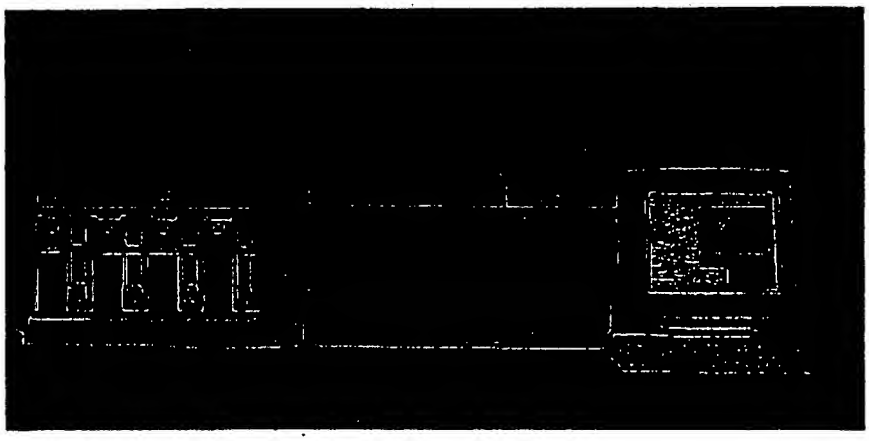

Figure 13.4 The GeneChip® Instrument System. From left to right, the hybridisation station, scanner and workstation. Image courtesy of Affymetrix. A colour version of this figure appears in the colour plate section

replicate samples prepared on different days. Pooling of reagents within an experiment, and analysing controls together with treated samples, will both reduce the variability within a given experiment.

Conduct of Experiment

At this stage some attention may be required for verifying and validating processes. For example, checking that the imaging instruments give consistent results across the slide, on repeat use and from day to day. If two imagers are used it is important to verify that the results from both machines are comparable. Some laboratories read fluorescence of one channel and then adjust the laser intensity of the second channel in order to obtain comparable readings. This is a method of normalising for the difference in intensities of fluorophores. It is important to be aware that this approach has a number of drawbacks. The arbitrary value of the second laser intensity setting will vary from experiment to experiment; thus comparisons of this channel cannot be made across experiments. Also the response of the fluorophore may not be linear across the laser intensity settings and this can lead to additional errors.

Another area for investigation prior to running the study itself is the image analysis component. Depending on the software used, the image analysis package may process the data to some extent, for example automatic background subtraction. Full understanding of the software is required so that it is clear at what point the data are 'raw', and the extent of inherent, inseparable manipulation. Effort may be required to determine the optimum settings for any software parameters.

As data are generated it is important to be aware of the data integrity – for example ensuring that all data are collected, so that there are no missing data that

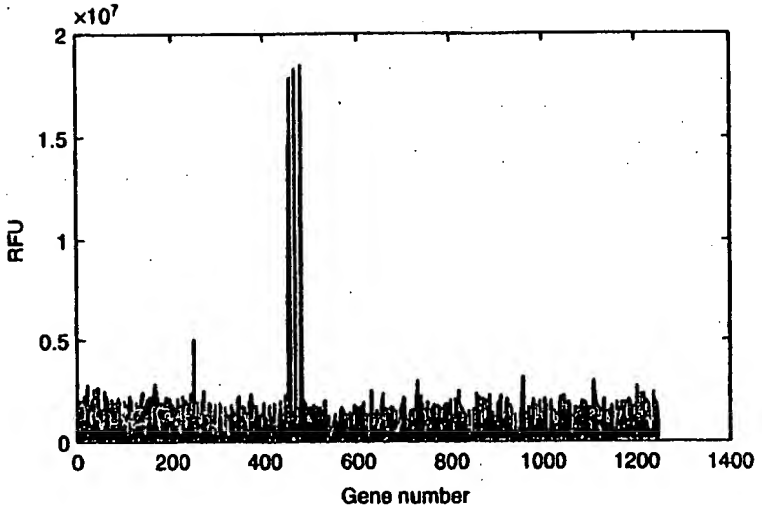

Figure 13.5 The relative fluorescence units (RFUs) of 1248 genes on seven microarray slides that were hybridised with cDNA made from the liver of a rat treated with acetaminophen. Note the gene outliers at approximately gene number 480

can complicate analysis later. The researcher may be intuitively aware of any spurious results and should be alert for anything extraordinary that could indicate problems, for example hybridisation intensities appearing inconsistent from sample to sample. Data analysis at this point can be a rapid indicator of dubious results. For example, Figure 13.5 shows a plot of the fluorescent intensities of 1248 genes that were hybridised with probe derived from acetaminophen-treated rat liver tissue. The data appear consistent, with the exception of peaks in intensity on one slide at around gene 4800. Further investigation of the microarray revealed a large artefact that had been missed by the image analysis process (Figure 13.6).

Raw Data Generation and Storage

One issue that arises when carrying out microarray analyses is how much data to store and in what form. For Good Laboratory Practice (GLP) purposes, often required in industry, storage of the raw data is necessary. This could be construed as the microarray image. Storing the image analysis results requires far less storage space and is easier to visualise, but it has the drawback that image analysis cannot be redone should superior software be available in the future. In reality, the methods used for microarrays are continually changing and the likelihood of revisiting old images on which the analysis has been performed, using outdated protocols is quite small.

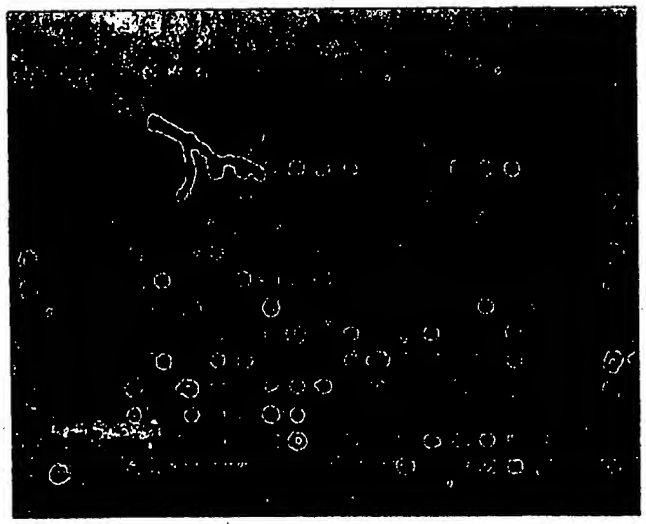

Figure 13.6 Portion of scanned image showing region where artefact occurred that caused very high signals, which were classed as outliers

Pre-processing

A number of pre-processing steps are often used in microarray analysis. These include filtering, log transformation, normalisation and background subtraction. Filtering may be used before or after transformation in order to extract data from preferred regions of interest, or in order to remove outliers (see the above example relating to image analysis artefact). One example of filtering is the removal of individual gene replicates that lie outside a given number (for example 5) of standard deviations from the mean. Alternatively, data points that lie in the top/bottom few percentiles (e.g. 0.1%) of the data can be removed. This method of removing outliers is also called 'trimming'. It is acceptable if there is a large volume of data where only a small proportion of data is removed and if the same method is applied consistently across all data. Care must be taken in the way in which this is carried out in order not to delete genuine data. For example, if one gene is consistently high or low in expression across replicates, then it is unlikely to be an outlier.

Another method of detecting outliers is to plot all genes (see the section 'Conduct of experiment' above) or to perform PCA analysis (see the section 'Multivariate analysis' below) to detect replicate outliers. The use of a PCA plot to detect outliers is shown in Figure 13.7.

Log transformation of data is accepted universally because the fluorescence data that are generated from microarrays tend to be skewed towards lower values. There are scientifically valid reasons why ratios of raw expression values should not be used (Nadon and Shoemaker, 2002). When using two-colour

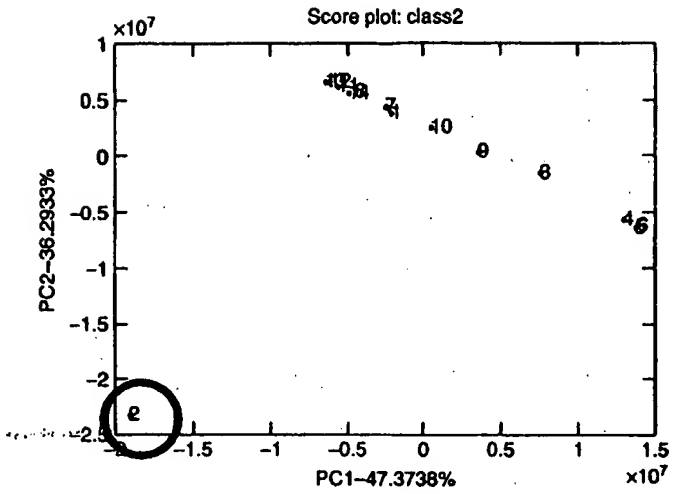

Figure 13.7 PCA plot of data used in Figure 13.4 showing one microarray (number 2) as an outlier

hybridisations it is common to express the ratio of treated to control as a logarithm in base 2 (Quackenbush, 2001). Thus genes up-regulated by a factor of 2 have a $\log_2$ (ratio) of 1, and genes down-regulated by a factor of 2 have a $\log_2$ (ratio) of -1.

Normalisation and background subtraction techniques are methods of data manipulation and their use is more subjective and often debated. The purpose of these techniques is to reduce the error (variability) that occurs between replicates and thus enable a comparison of data across samples.

The theory behind background subtraction is that during hybridisation there will be non-specific binding to the slide. This will effectively 'darken' the image and give falsely high readings of fluorescent intensity. Correcting for the non-specific hybridisation should reduce error due to background staining. Background subtraction often occurs automatically in microarray image analysis packages. The software may circle the spot of interest and use the region beyond the periphery as the measurement of background. In cases of uneven hybridisation this method enables locally high background to be subtracted on a regional basis. One criticism of this approach is that the slide surface beyond the periphery is not similar, in chemical terms, to that where the nucleic acid has been deposited, and therefore cannot act as a real control for non-specific binding. A more accurate measurement of non-specific binding can be gained from using a region where spotting chemicals have been deposited, but no target is present. This concept is the basis of a method using local 'blank spots' (Wu *et al.*, 2001).

A number of methods exist for normalisation of data. These include normalising to total signal or to a 'known' spot or gene, standardisation, or proprietary

methods. Normalising to total signal is the simplest approach, whereby the gene intensity is expressed as a percentage or proportion of the signal intensity for the entire array. This method works best when the total intensities for the microarrays are similar and the number of changes is small compared with the number of genes. However, we often find, when using arrays of around 1000 genes, that pathological disease can up-regulate a large number of genes simultaneously. In this case, when total signal normalisation is applied, highly up-regulated genes will appear less up-regulated and genes that do not change from the control will appear down-regulated.

Normalisation to a control value is a more popular technique. A control value can be obtained from using a gene known to remain constant under the conditions of the experiment. DeRisi *et al.* (1997) used a panel of 90 housekeeping genes for normalisation, but found considerable variation in their gene expression. Unfortunately it is very difficult to know with certainty that a gene will not change and there is evidence to suggest that so-called 'housekeeping genes' are variable (Savonet *et al.*, 1997). Other control genes can be derived from an alternative species; these should not be expected to hybridise. We have used yeast and Arabidopsis genes as negative controls for hybridisations of rat tissues. No orthologs were known to the genes selected; however, in most cases non-specific binding occurred.

If two or more microarray replicates appear to be different, but they are expected to be the same, then they can be standardised. An example of this might occur if the total intensity of one microarray is greater than another, but the genes are proportionally equally up- or down-regulated. If the microarray sample spot data are assumed to be drawn from a normal distribution, then the 'z-transform' can be used. This requires that the mean and the standard deviation of the intensity values for each microarray are determined. The mean is subtracted from each individual gene value and the remainder is divided by the standard deviation. The intensity values from each microarray will then have the same mean and standard deviation. This has the advantage of facilitating comparison of microarrays with different dynamic ranges as well as total intensities. If the data are not normally distributed, then alternative non-parametric methods can be used, such as normalising to the median.

Univariate Analysis

Univariate methods of analysis involve examining one variable, or gene, at a time. This can be a very laborious task when examining a large volume of data, yet it is the preferred method of biologists. The simplest technique is to compare control and treated values and express the result as a 'fold-change' ratio. Typically, when examining small volumes of data, fold changes greater than 2 and less than 0.5 are considered meaningful (Quackenbush, 2001). This cut-off is essentially arbitrary and has the distinct drawback that microarray data

are not homoscedastic; that is, there is more variation about the mean at low values than there is at high values (Draghici, 2002).

A second method for finding up- or down-regulated genes uses the standard deviation (SD) of the replicate gene data. Thus if changes greater than, say, 2 SD from the log mean ratio are considerably greater than changes associated with 'noise', then they are considered significant. This technique means that when looking at a large number of genes that are normally distributed there will be up- and down-regulated genes, regardless of whether there are (biological) changes (Draghici, 2002).

Rather than using arbitrary cut-off values it is far more meaningful to express the fold-change in terms of either confidence intervals, or a '*p*-value' (that is, the probability of the value occurring by chance). Thus, a fold-change of 1.1 may be associated with a *p*-value of 0.001 and thus the probability that the gene is *not* up-regulated is 1 in 1000. Naturally, such small fold-changes may then be queried in terms of biological significance. One must then ask the question: Are large fold-changes more important (biologically) than small ones? Simple calculation of *p*-values for two data sets can be obtained using *t*-test functions in standard spreadsheet software. A number of replicates are necessary for this approach, and the data must be normally distributed. We have recently developed a method for calculating *p*-values for fold-changes that is not influenced by the distribution of the data or outliers and applied it to TaqMan™ and microarray data. Other complex and computationally intensive methods for calculating *p*-values are described in Draghici (2002) and Nadon and Shoemaker (2002).

Multivariate Analysis

Multivariate analysis of gene expression data is becoming increasingly popular in the microarray community and in other biological domains where large volumes of data are generated. Multivariate analysis methods include principal component analysis (PCA), factor analysis, multivariate analysis of variance (MANOVA) and cluster analysis. Currently, cluster analysis is the most widely-used method in the microarray community but PCA is growing in popularity (Crescenzi and Giuliani, 2001; Konu *et al.*, 2001).

Quackenbush (2001) provides a good review of clustering tools that is rather unique in the regard that different clustering algorithms and linkage methods are presented. Clustering methods are unsupervised, and they are powerful tools for gaining insight into huge data sets. They enable the data to be partitioned in order to facilitate interpretation; however, they do suffer from subjectivity. This is because the user selects various parameters, such as the algorithm used, linkage type, distance metric and, sometimes, cluster size. Whatever the data, clusters will always be identified, thus there is also a tendency to over-interpret the data – trying to attach meaning to clusters that may have no biological relevance.

Software available for cluster analysis includes Cluster, the output of which is viewed in Treeview; both available from `http://rana.stanford.edu/software`. This tool is particularly useful for clustering genes to identify genes that are co-regulated.

The PCA is a visualisation tool that enables complex, high-dimensional data to be represented in two or three dimensions. It facilitates identification of groups of similar data, thus enabling inferences to be made about the samples.

An example is shown in Figure 13.8. The figure shows the gene data for one sample (control rat liver) that was hybridised to seven microarrays according to the method used in Wildsmith *et al.* (2001). Each microarray contained two replicate gene sets; thus there were 14 replicate gene sets in total. The gene sets comprised 1248 genes (with controls). All the data (14 × 1248 data points) was input into the analysis and the PCA plot displays the 14 replicates individually. The two axes are principal components 1 and 2. Principal component 1 (PC1) accounts for 65.5% of the variation in the data, whereas PC2 represents only 13%. This means that the model accounts for 78.5% of variation in the data.

The first principal component (PC1) accounts for as much as possible of the variation in the original data and subsequent components (e.g. PC2) are of decreasing importance. Thus, samples 8 and 9 are very different from samples 1 and 2. In terms of interpreting the PCA plot, it is immediately clear that there are three or four distinct clusters of data. These are marked by circles. Datapoints tend to cluster in pairs; for example replicates 1 and 2, 3 and 4, 5 and 6, etc. These are the duplicate gene sets on the same microarray. This indicates that the variation within the microarray is lower than the variation between

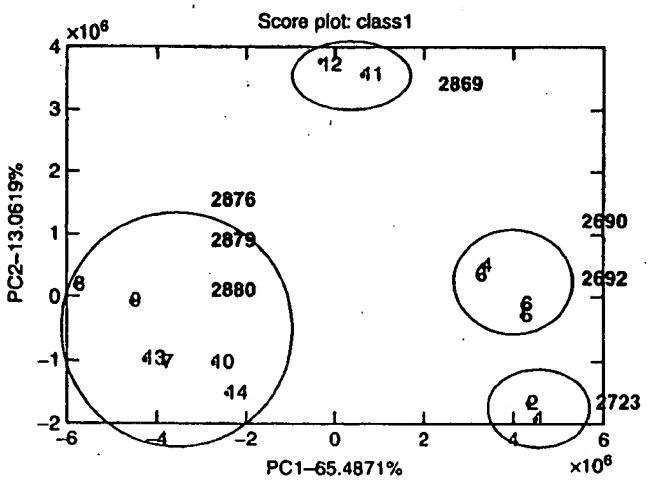

Figure 13.8 PCA plot of the microarray results from seven slides (2690, 2692, 2723, 2869, 2876, 2879, 2880), each with two replicate spot sets (labelled 1–14), after hybridisation with control rat liver. See text for explanation

replicate microarrays. However, given that the same sample is applied to all the microarrays, we must ask why we get further separation of the replicates. The answer lies in the associated data. The four-digit numbers associated with the clusters are the microarray slide numbers that indicate when they were printed. Numbers that are more similar seem to be more closely related, and thus we can hypothesise that there were some differences between the slides, such as differences in slide backgrounds or changes during the printing process or change-over of batches, between slides in the 2600–2700 region and the 2800s.

The PCA provides a clearer overview of the data than does cluster analysis. It is a rapid method for gaining an insight into the results, in particular where biological meaning can be attached to the components (Crescenzi *et al.*, 2001). There are a number of packages for multivariate data analysis, including SIMCA-P (Umetrics, AB, Umea, Sweden) and The Unscrambler (Camo ASA, Norway), both of which are useful for PCA.

Other tools for data visualisation include software packages such as Spotfire.net (Spotfire Inc., Cambridge, MA, USA) and GeneSpring (Silicon Graphics, San Carlos, CA, USA). Spotfire is particularly useful for visualisation of multidimensional data and for visualisation of temporal data. It is possible to use these tools to identify genes that are co-ordinately expressed over time.

Biological Interpretation

After developing a sound experimental strategy, ensuring that the results are statistically valid, and after analysis of the data, it is down to the biologist to assemble the pieces of information that have been obtained. This intertwined information may include unexpected results that are contradictory to intuition or to published literature. One way to untangle the data is to map the relevant genes onto existing pathways and known functions. The Kyoto Encyclopedia of Genes and Genomes (KEGG), available at `http://www.genome.ad.jp/kegg/`, is a useful source of information, especially where the gene products are enzymes. It enables visualisation of the position of up- or down-regulated genes in metabolic pathways.

The gene expression data obtained may differ from protein expression data, or information on gene product activity or location. When initiating a study it is useful to consider additional endpoints that can assist in the interpretation of the data. For *in vitro* studies, these might include cytotoxicity endpoints, metabolites, key signalling molecules or perhaps protein expression. Waring *et al.* (2001) used tetrazolium dye reduction (MTT) as a measure of hepatocyte cell viability for their studies of gene expression in response to hepatotoxic insult. For *in vivo* studies, expression information on the tissue of interest could be supported by pathology, histology and blood chemistry measurements. Gene expression results could be confirmed by *in situ* hybridisations or protein activity assays.

13.4 Recent examples of microarray applications

One area of rapid progress using microarray technology is the increased understanding of cancer. Molecular pathologists are subgrouping cancers of tissues such as blood, skin and breast, based on differential gene expression patterns. For example, within a small group of breast cancer tissue samples, Perou *et al.* (2000) distinguished two broad subgroups representing those expressing or alternatively lacking expression of the oestrogen receptor-α gene. The work was not conclusive, but never has progress in this field been so rapid when compared with the previous methods of gene identification.

Another example of the impact of this technology is in the identification of two biomarkers for prostate cancer, namely hepsin and PIM1 (Dhanasekaran *et al.*, 2001).

Microarray technology has also accelerated the understanding of the molecular events surrounding pulmonary fibrosis. Specifically, two distinct clusters of genes associated with inflammation and fibrosis have been identified in a disease where, for years, the pathogenesis and treatment have remained unknown (Katsuma *et al.*, 2001).

13.5 Conclusions

Important factors in gene expression experiments include sensitivity, precision and reproducibility in the measurement of specific mRNA sequences (Schmittgen *et al.*, 2000). These quality metrics can be maximised by using, or fabricating, high-quality microarrays, and by optimising each step of the microarray process. From conception to conclusion it is important to bear in mind the original hypothesis.

Having considered the complexity of the microarray experiment, the value obtained from a meticulously designed experiment should not be underestimated. As the number of high-quality gene expression studies increases, we hope that the literature will contain increasingly detailed information that will help interpret complex gene expression changes, and thus elucidate the mechanisms of disease.

13.6 Acknowledgements

The authors express their thanks to W. Wu for Figures 13.7 and 13.8, and to W. Wu, S. Roberts and P. Lane for useful comments and discussions.

13.7 References

Bowtell DL (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nature Genet.* **21**: 25–32.

Box GEP, Hunter WG and Hunter JS (1978) Chapter 7, in: *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. John Wiley and Sons, New York.

Cheung VG, Morley M, Aguilar F *et al.* (1999) Making and reading microarrays. *Nature Genet.* **21**: 15–19.

Crescenzi M and Giuliani A (2001) The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. *FEBS Lett.* **507**: 114–18.

DeRisi JL, Iyer VR and Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–6.

Dhanasekaran SM, Barrette TR, Ghosh D *et al.* (2001) Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**: 822–4.

Draghici S (2002) Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov. Today*, in press.

Fernandez PM, Pluta LJ, Fransson-Steen R *et al.* (1997) Reverse transcription-polymerase chain reaction-based methodology to quantify differential gene expression directly from microdissected regions of frozen tissue sections. *Mol. Carcinog.* **20**: 317–26.

Golub TR, Slonim DK, Tamayo P *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**: 531–7.

Hegde P, Qi R, Abernathy K *et al.* (2000) A concise guide to cDNA microarray analysis. *Biotechniques* **29**(3): 548–62.

Katsuma S, Nishi K, Tanigawara K *et al.* (2001) Molecular monitoring of bleomycin-induced pulmonary fibrosis by cDNA microarray-based gene expression profiling. *Biochem. Biophys. Res. Commun.* **288**: 747–51.

Konu O, Kane JK, Barrett T *et al.* (2001) Region-specific transcriptional response to chronic nicotine in rat brain. *Brain Res.* **909**: 194–203.

Lee TML, Kuo FC, Whitmore GA *et al.* (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridisations. *Proc. Natl. Acad. Sci. USA* **97**(18): 9834–9.

Nadon R and Shoemaker J (2002) Statistical issues with microarrays: processing and analysis. *Trends Genet.* **18**(5): 265–71.

Nuwaysir EF, Bittner M, Trent J *et al.* (1999) Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* **24**: 153–9.

Perou CM, Sorlie T, Eisen MB *et al.* (2000) Molecular portraits of human breast tumours. *Nature* **406**: 747–52.

Quackenbush J (2001) Computational analysis of microarray data. *Nature Rev. Genet.* **2**: 418–27.

Savonet V, Maenhaut C, Miot F *et al.* (1997) Pitfalls in the use of several 'housekeeping' genes as standards for quantitation of mRNA: the example of thyroid cells. *Anal. Biochem.* **247**: 165–7.

Schmittgen TD, Zakrajsek BA, Mills AG *et al.* (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal. Biochem.* **285**: 194–204.

Shalon D, Smith SJ and Brown PO (1996) A DNA microarray system for analysing complex DNA samples using two-colour fluorescent probe hybridisation. *Genome Res.* **6**: 639–45.

Taniguchi M, Miura K, Iwao H *et al.* (2001) Quantitative assessment of DNA microarrays – comparison with northern blot analyses. *Genomics* **71**: 34–9.

Van Hal NLW, Vorst O, Van Houwelingen MML *et al.* (2000) The application of DNA microarrays in gene expression analysis. *J. Biotech.* **78**: 271–80.

Waring JF, Ciurlionis R, Jolly RA *et al.* (2001) Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol. Lett.* **120**: 359–68.

Wildsmith SE and Elcock FJ (2001) Microarrays under the microscope. *J. Mol. Pathol.* **54**: 8–16.

Wildsmith SE, Archer GEB, Winkley AJ *et al.* (2001) Maximisation of signal associated with cDNA microarrays. *BioTechniques* **30**: 202–8.

Wu W, Wildsmith SE, Winkley AJ *et al.* (2001) Chemometric strategies for normalisation of microarray data. *Anal. Chim. Acta.* **446**: 449–64.

13.8 Further reading

Chatfield C and Collins AJ (1980) *Introduction to Multivariate Analysis.* Chapman & Hall, London.

Heck DE, Roy A and Laskin JD (2001) *Nucleic Acid Microarray Technology for Toxicology: Promise and Practicalities. Biological Reactive Intermediates VI.* Kluwer Academic Press.

Schena M (ed.) (2001) *Microarray Biochip Technology.* Eaton Publishing.

13.9 Useful websites

Protocols:

`http://cmgm.stanford.edu/pbrown`
`http://cmgm.stanford.edu/pbrown/protocols.html`

Image analysis software:

`http://www.genome.ad.jp/kegg/`
`http://rsb.info.nih.gov/nih-image`
`http://rana.stanford.edu/software`
`http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html`
`http://rana.Stanford.EDU/software`
`http//www.tigr.org/softlab`

Oncogene (2006) 25, 2628–2635

www.nature.com/onc

ORIGINAL ARTICLE

Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach

R Li[1], H Wang[1], BN Bekele[2], Z Yin[3], NP Caraway[1], RL Katz[1], SA Stass[4] and F Jiang[4]

[1]*Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA;* [2]*Department of Biostatistics & Applied Mathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA;* [3]*Department of Internal Medicine, The University of Texas Houston Medical School, Houston, TX, USA and* [4]*Department of Pathology, The University of Maryland School of Medicine, Baltimore, MD, USA*

Amplification and overexpression of putative oncogenes confer growth advantages for tumor development. We used a functional genomic approach that integrated simultaneous genomic and transcript microarray, proteomics, and tissue microarray analyses to directly identify putative oncogenes in lung adenocarcinoma. We first identified 183 genes with increases in both genomic copy number and transcript in six lung adenocarcinoma cell lines. Next, we used two-dimensional polyacrylamide gel electrophoresis and mass spectrometry to identify 42 proteins that were overexpressed in the cancer cells relative to normal cells. Comparing the 183 genes with the 42 proteins, we identified four genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – in which elevated protein expression correlated with both increased DNA copy number and increased transcript levels (all $r > 0.84$, two-sided $P < 0.05$). These findings were validated by Southern, Northern, and Western blotting. Specific inhibition of EEF1A2 and KCIP-1 expression with siRNA in the four cell lines tested suppressed proliferation and induced apoptosis. Parallel fluorescence *in situ* hybridization and immunohistochemical analyses of *EEF1A2* and *KCIP-1* in tissue microarrays from patients with lung adenocarcinoma showed that gene amplification was associated with high protein expression for both genes and that protein overexpression was related to tumor grade, disease stage, Ki-67 expression, and a shorter survival of patients. The amplification of *EEF1A2* and *KCIP-1* and the presence of overexpressed protein in tumor samples strongly suggest that these genes could be oncogenes and hence potential targets for diagnosis and therapy in lung adenocarcinoma.
Oncogene (2006) **25,** 2628–2635. doi:10.1038/sj.onc.1209289; published online 12 December 2005

Keywords: lung cancer; microarrays; proteomics; tissue microarray

Introduction

In lung adenocarcinoma, as in other types of cancer, gene amplification and the consequent overexpression of the amplified oncogene play an important role in the development of tumors, because their overexpression confers a growth advantage. The ability to identify putative oncogenes that are activated during tumorigenesis could facilitate the choice of molecular genetic targets for diagnosis and therapy of the disease. This concept has been exemplified by *HER-2*, which was first found to be amplified in neuroblastomas and subsequently shown to be associated with poor prognosis in breast cancer (Ross and Fletcher, 1999). Now, *HER-2* aberrations are used as a predictor of response to therapy, and treatment of HER-2-positive breast cancer with the monoclonal anti-HER-2 antibody trastuzumab has been shown to improve prognosis (Ross and Fletcher, 1999). Emerging evidence of common amplicons in lung adenocarcinomas (Luk *et al.*, 2001; Jiang *et al.*, 2004; Tonon *et al.*, 2005) suggests that additional oncogenes remain to be identified; however, conventional techniques are ineffective in pinpointing such oncogenes. Parallel measurement of DNA copy number and mRNA levels in cDNA microarrays permits changes in copy number to be compared with transcription levels on a gene-by-gene basis to generate lists of candidate genes within the defining amplicons (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, use of transcript patterns does not allow assessment of the expression of protein products or identification of proto-oncogenes. Another approach, identifying differentially expressed proteins by proteomic analysis and then comparing the proteins present with mRNA expression in cDNA microarrays from the same specimens, can clarify the extent to which changes in transcript patterns reflect changes in their cognate proteins and post-transcriptional mechanisms (Chen *et al.*, 2002), but this approach cannot be used to identify oncogenes driven by extensive increases of their gene copy number. Moreover, using individual microarrays or proteomic approaches alone cannot distinguish the cancer-driving oncogenes that directly propel tumor progression from the larger number of passenger genes that may be concurrently over-represented but are not biologically relevant in tumor development.

Correspondence: Assistant Professor F Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th floor, Baltimore, MD 21201-1192, USA.
E-mail: fjiang@som.umaryland.edu
Received 27 July 2005; revised 19 October 2005; accepted 27 October 2005; published online 12 December 2005

In this study, we used a comprehensive approach that integrated simultaneous comparative genomic hybridization (CGH) and transcript microarray with proteomic analyses of six lung adenocarcinoma cell lines. We directly and specifically identified four putative oncogenes that could have been activated through amplification and consequent elevation of transcript expression. We used small interfering RNA (siRNA) to inhibit the expression of two of these four genes in the lung cancer cell lines, which further implicated them in oncogenesis. We then explored the clinical significance of these findings by assessing the expression of these two genes in tissue microarrays of human lung cancer specimens. Our findings underscore the power of integrated functional genomic analyses for identifying putative oncogenes in tumorigenesis; such activated genes could be useful as targets for diagnosis or therapy in lung cancer.

Results

Simultaneous global genomic and transcript analyses identify 183 genes with increases in genomic copy numbers and transcript expression levels

To identify genes in which increased DNA copy number might contribute to increased transcript in lung adenocarcinomas, first we used CGH with microarrays of six lung adenocarcinoma cell lines. We identified 587 genes showing increases in DNA copy number across all six cell lines (Supplementary Table 1S), which were distributed as 90 amplicons on all chromosomes except for chromosomes 13 and Y (Supplementary Table 2S). A subsequent transcript test with the identical arrays of the same cell lines revealed 275 genes that showed increased mRNA levels (Supplementary Table 3S). Using random permutation tests across all cancer cell lines, we identified 183 genes (31%) that showed elevated transcript levels from the 587 genes that were over-represented in the genome (Table 1), suggesting that elevated transcript levels of the 183 genes may reflect their genomic over-representation in the cancer cells. These findings are consistent with previous reports linking genomic changes with altered transcript patterns in breast cancer (Hyman *et al*., 2002; Pollack *et al*., 2002). However, our finding that only 31% of the genes showing increased DNA copy numbers had cognate increases in transcript expression in lung adenocarcinomas is different from the overall rates of 40–60% reported for breast cancer (Hyman *et al*., 2002; Pollack *et al*., 2002). This discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma.

Proteomic analyses identify four genes for which protein abundance was associated with increases in the cognate gene and transcript levels

Analysis of transcript patterns is insufficient for understanding the expression of protein products and the effect of genomic over-representation on the expression of their cognate proteins. To extend these findings beyond genomic over-representation to expression of the protein products of those genes, we next assessed protein expression in the same cell lines by two-dimensional polyacrylamide gel electrophoresis (PAGE) and found that 42 different proteins, representing 42 individual genes, were significantly increased in the cancer cell lines (Table 2; Supplementary Figures 1S and 2S). Some of these proteins were identified as having multiple isoforms, and all individual isoforms exhibited increases in expression ranging from 4.6 to 12.8 times their expression in normal lung tissue cells. In comparing protein level of the 42 genes with changes in their cognate genomic and mRNA expression from the global microarray analyses, we found that four (9.5%) of those 42 genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – showed statistically significant correlations between elevated protein expression and increases in both copy number and mRNA expression (all $r > 0.84$; $P < 0.05$) (Table 2) in the cancer cell lines. These findings imply that the abundance of these four proteins is attributable to the amplification and consequent elevated transcription of their cognate genes.

Validation of copy number, transcript, and protein expression of PRDX1, EEF1A2, CALR, and KCIP-1 in lung cancer cell lines

To confirm our findings from the high-throughput analyses, we next used Southern, Northern, and Western blotting to assess DNA, RNA, and protein levels for the four genes identified in the six cell lines. For comparison, we arbitrarily chose one gene, *NFKBI*, in which an increase in protein level did not correlate with genetic changes. Overall, we found excellent concordance between the CGH microarray and Southern blotting analyses, transcript array and Northern blotting analyses, and proteomic and Western blotting analyses for all five genes (Figure 1). For example, *KCIP-1* showed fivefold amplification in five of the six cancer cell lines, whereas NFKBI showed no such increase in any of the cell lines. As for transcript expression, Northern blotting of *EEF1A2* showed high expression in five of the six cancer cell lines; again, levels of *NFKBI* transcript were not increased in any cancer cell line as compared with normal bronchial epithelial cells. The results of Western blotting were also consistent with the results of the proteomic experiments; for example, five of the cancer cell lines exhibited strong protein bands for PRDX1 as compared with normal cells. These findings provide strong support for the validity of the results derived from the high-throughput techniques in this study.

These parallel analyses also revealed close correlations in the extent of changes in gene copies, transcript, and protein of each of the four genes in the cancer cell lines. For example, in the five cancer cell lines that showed at least fourfold increases in *EEF1A2* copy number, expression of transcript and protein was also increased by at least a factor of four as well (relative to their expression in normal cells) (Supplementary Figure 3S). The protein abundance of the four genes showing

Table 1 List of 183 genes with statistically significant correlation (0.05) between genomic copy number and transcript level

Gene symbol	*Chro.*	*Distance from p arm of each chromosome (Mb)*	α
ENO1	1	8.5	0.0085
DDOST	1	20.1	0.0111
SFN	1	26.4	0.0113
MLP	1	32.2	0.0114
AKR1A1	1	45.4	0.0128
PRDX1	1	45.4	0.0122
UQCRH	1	46.2	0.0125
RPL7	1	96.4	0.0127
COL11A1	1	102.6	0.0129
MCL1	1	147.3	0.0222
PSMB4	1	148.1	0.0131
JTB	1	150.7	0.0134
RPS27	1	150.7	0.0135
HAX1	1	151	0.0266
MUC1	1	151.9	0.0143
CCT3	1	153.1	0.0167
CRABP2	1	153.4	0.0148
TKT	1	159.3	0.0152
ATP1B1	1	165.8	0.0234
CHIT1	1	199.7	0.0154
SNRPE	1	200.2	0.0165
YWHAQ	2	9.6	0.0159
ODC1	2	10.60	0.0119
RPL31	2	101.20	0.0161
BENE	2	110.40	0.0169
STAT1	2	191.80	0.0175
HSPD1	2	198.30	0.0277
HSPE1	2	198.30	0.0185
RPL37A	2	217.30	0.0388
IGFBP2	2	217.50	0.0189
RPS7	2	3.30	0.0193
RABIA	2	65.30	0.0204
IGKC	2	89.00	0.0285
LTF	3	46.3	0.0455
PFN2	3	151	0.0207
KPNA4	3	161.5	0.0211
S100P	4	6.7	0.1122
UGDH	4	39.3	0.0215
UCHL1	4	41.1	0.0222
SPP1	4	89.3	0.0227
TRIM2	4	154.7	0.0231
FGB	4	156	0.0235
FGG	4	156	0.0441
SDHA	5	0.251	0.0243
PDCD6	5	0.305	0.0245
CCT5	5	10.3	0.0446
PTPRF	5	14.2	0.0248
RPL37	5	40.8	0.0251
ENC1	5	74	0.0336
QP-C	5	132.2	0.0466
SPINK1	5	147.2	0.0256
CANX	5	179.2	0.0263
SOX4	6	21.7	0.0321
HDGF	6	22.6	0.0362
RPS10	6	34.6	0.0177
RPL10A	6	35.4	0.0369
VEGF	6	43.7	0.0372
OSF-2	6	45.4	0.0173
FSCN1	7	5.3	0.0378
CYCS	7	24.9	0.0381
CBX3	7	25.9	0.0289
IGFBP3	7	45.7	0.0389
CLDN4	7	72.7	0.0403
HSPB1	7	75.5	0.0433
CALR	7	92.7	0.0425
COL1A2	7	93.6	0.0457
ATP5J2	7	98.7	0.0475
AKR1B10	7	133.6	0.0481
RPS20	8	56.7	0.0482
TCEB1	8	74.6	0.0486
LAPTM4B	8	98.5	0.0497
RPL30	8	98.7	0.0054
KCIP-1	8	101.6	0.0093
PABPC1	8	101.78	0.0119
EEF1D	8	144.4	0.0121
TSTA3	8	144.5	0.0122
RPL8	8	145.6	0.0128
TRA1	9	117.1	0.0136
RPL35	9	121.1	0.0133
HSPA5	9	121.5	0.0135
LCN2	9	124.4	0.0137
DPP7	9	133.4	0.0139
PFKP	10	3.2	0.0223
AKR1C1	10	5.1	0.0146
PLAU	10	75.6	0.0356
DSP	10	76.7	0.0289
TALDO1	11	0.434	0.0143
SLC22A1L	11	2.9	0.0151
TSSC3	11	2.9	0.0611
RPL27A	11	8.7	0.0156
ST5	11	8.8	0.0162
LDHA	11	18.5	0.0168
MDK	11	46.4	0.0162
DOC-1R	11	67.5	0.0167
MMP12	11	102.8	0.0177
HYOU1	11	118.9	0.0183
SCNN1A	12	6.3	0.0185
LDHB	12	21.7	0.0193
KRT7	12	52.3	0.0196
KRT5	12	52.6	0.0197
KRT6E	12	52.6	0.0201
ERBB3	12	56.2	0.0212
NACA	12	56.8	0.0218
TM4SF3	12	71.2	0.0401
NTS	12	86.2	0.0215
ASCL1	12	103.3	0.0219
TXNRD1	12	104.6	0.0223
CKAP4	12	106.6	0.0124
COX6A1	12	120.7	0.0435
BGN	12	122.5	0.0235
RAN	12	129.88	0.0238
RPL36A	14	48.1	0.0243
PGD	14	50.7	0.0248
THBS2	15	37.5	0.0251
TRAF4	15	38.3	0.0253
SPINT1	15	38.7	0.0254
RPL17	15	45.26	0.0411
PKM2	15	70.1	0.0258
IDH2	15	88.2	0.0211
RPL23A	16	0.377	0.0264
MSLN	16	0.753	0.0366
UBE2I	16	1.3	0.0271
RPS2	16	1.95	0.0281
CLDN9	16	3.1	0.0329
ARL6IP	16	18.7	0.0412
EIF3S8	16	28.3	0.0336
TUFM	16	28.9	0.0377
ALDOA	16	30.1	0.038
NME4	16	53.6	0.0381
GPR56	16	57.4	0.0386
CDH1	16	68.5	0.0289
NQO1	16	69.5	0.0396
SLC7A5	16	87.6	0.0397
APRT	16	88.6	0.0411
GALNS	16	88.6	0.0255
RPL13	16	89.3	0.0431
MCP	17	32.4	0.0465

Table 1 (*continued*)

Gene symbol	*Chro.*	*Distance from p arm of each chromosome (Mb)*	α
ERBB2	17	35.11	0.0483
JUP	17	39.8	0.0495
CRF	17	40.39	0.0505
RPL27	17	41.1	0.0046
NME1	17	46.59	0.0082
COL1A1	17	48.6	0.0108
ABCC3	17	49.1	0.0326
NME2	17	49.6	0.0111
RPL38	17	72.7	0.0117
SMT3H2	17	73.6	0.0119
SYNGR2	17	76.6	0.0122
LGALS3BP	17	77.4	0.0127
P4HB	17	80.3	0.0126
PPAP2C	19	0.221	0.0228
GPI	19	39.55	0.0145
HPN	19	40.2	0.0129
ZNF146	19	41.4	0.0131
SPINT2	19	43.4	0.0238
PSMD8	19	43.5	0.0132
YIF1P	19	43.5	0.0135
RPS16	19	44.6	0.0144
CEACAM5	19	46.9	0.0145
CEACAM6	19	46.9	0.0143
GIPR	19	50.8	0.0259
SNRPD2	19	50.9	0.0413
KDELR1	19	53.6	0.0152
RPL28	19	60.6	0.0156
RPS5	19	63.6	0.0267
TRIM28	19	63.7	0.0158
DAP	20	35.6	0.0166
TOP1	20	40.3	0.0172
UBE2C	20	45.1	0.0174
RPS21	20	61.6	0.0268
EEF1A2	20	62.8	0.0185
TFF3	21	42.6	0.0186
TFF1	21	42.7	0.0192
CSTB	21	44.1	0.0201
MIF	22	22.6	0.0202
XBP1	22	27.5	0.0204
PRDX4	X	22.9	0.0198
SYN1	X	46.3	0.0204
TIMP1	X	46.3	0.0209
PLP2	X	47.8	0.0212
MAGED1	X	50.3	0.0331
RPS4X	X	71	0.0124
SSR4	X	152.6	0.0232

corresponding increases in both DNA copy number and mRNA provides further evidence that these could be oncogenes, the activation of which is reflected by genomic amplification and consequent increases in transcript level in lung adenocarcinoma cell lines.

Specific inhibition of EEF1A2 and KCIP-1 expression by siRNAs led to decreased cell proliferation and induction of apoptosis

To further prove the oncogenic function of the identified genes in lung tumorigenesis, we used siRNAs to inhibit the endogenous expression of EEF1A2 and KCIP-1 protein in four lung cancer cell lines (H1563, H229, H522, and SK-LU). Transfection of the cancer cells with specific siRNAs reduced the level of EEF1A2 and KCIP-1 protein by 70–90% 48 h after transfection (Supplementary Figure 4S). In contrast, EEF1A2 and KCIP-1 protein levels remained unchanged in mock-treated control cells and in cells transfected with a scrambled siRNA sequence. At 48 h after siRNA transfection, the percentage of proliferation of the transfected cancer cells was reduced to 15–30% as compared with 91–100% of cell proliferation of the same cell lines treated with PBS or scrambled siRNA (Supplementary Figure 5S). Apoptosis of siRNA-transfected cells was 27–34%, whereas only 4% of the same cell lines treated with PBS or scrambled siRNA showed apoptosis. These results strongly support an oncogenic role for the identified genes in lung cancer and confirm their potential usefulness as therapeutic targets for the disease.

Amplification and protein expression of KCIP-1 and EEF1A2 in lung tissue

To further validate these findings and to assess the possible clinical significance of the four potential putative oncogenes identified from the cell lines, we first applied fluorescence *in situ* hybridization and immunohistochemical analysis, in parallel, to commercially available human lung tissue microarrays (Ambion, Austin, TX, USA) to evaluate the status of two of these four genes in lung cancer tissue specimens. (Commercially available antibodies to PRDX1 or CALR were not suitable for use in immunohistochemical analysis when this report was written.) Overexpression of KCIP-1 and EEF1A2 protein in the tumors was concordant with amplification of the corresponding genes ($P=0.0003$ for KCIP-1 and $P=0.0011$ for EEF1A2). For example, 16 (35%) of the 46 lung adenocarcinomas in the microarray showed amplification of *KCIP-1*, and strong cytoplasmic staining for KCIP-1 protein was seen in 18 tumors (39%) (Figure 2). We next examined whether overexpression of these genes was associated with increased cell proliferation by analysing Ki-67 expression in contiguous sections of the tissue microarrays. Positive Ki-67 expression was found to correlate with positive expression of both KCIP-1 ($P=0.02$) and EEF1A2 ($P=0.01$). To extend these findings, we then studied 11 tissue microarray blocks comprising normal and tumor tissue specimens from 113 patients with pathologic stage I non-small-cell lung cancer who had undergone curative surgery (Wang *et al.*, 2005). Immunohistochemical analysis showed that EEF1A2 was expressed in 32 cases (28%) and KCIP-1 in 29 cases (26%). Univariate and multivariate Cox proportional hazards models were used to detect possible associations between EEF1A2 and KCIP-1 expression and clinicopathologic variables. Expression of EEF1A2 or KCIP-1 was associated with short overall survival time ($P=0.0012$ for EEF1A2 and $P=0.0026$ for KCIP-1) (Supplementary Figure 6S). Age at diagnosis, histologic type of cancer, degree of tumor differentiation, and smoking history were not associated with survival time.

Although only two genes were validated in the lung tissue microarrays (because available antibodies to the other two genes were not suitable for use in

Table 2 Proteins showing significant overexpression in cancer cell lines relative to those in normal bronchial epithelial cell lines and their correlation coefficients with increased DNA copy number or mRNA values[a]

Acc. no.	*Gene ID*	*Gene*	*Mw/pI*	*Description*	*r with genomic copy changes*[b]	*r with mRNA changes*[b]
Q06830	5052	PRDX1	48.4/5.4	Peroxiredoxin 1	**0.92364**	**0.91892**
Q05639	1917	EEF1A2	50.5/5.7	Eukaryotic translation elongation factor 1 alpha 2	**0.90218**	**0.89456**
P27797	811	CALR	61/5.5	Calreticulin	**0.84128**	**0.86434**
P63104	7534	KCIP-1	27/6.5	Tyrosine 3-monooxygenase activation protein, zeta	**0.84467**	**0.85499**
P07237	5034	P4HB	54/6.2	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	**0.91884**	0.76786
Q04695	3872	KRT17	48.0/4.9	Keratin 17	0.00236	**0.86892**
P09211	2950	GSTP1	23.2/4.7	Glutathione *S*-transferase pi	**0.84218**	0.69456
P17936	3486	IGFBP-3	31.6/5.8	Insulin-like growth-factor binding protein 3	0.06412	0.16434
P26641	1937	EEF1G	50/6.4	Eukaryotic translation elongation factor 1 gamma	0.00446	**0.85549**
P08727	3880	KRT19	44.1/5.2	Keratin 19	−0.04884	**0.86786**
P04792	3315	HSPB1	22/6.5	Heat shock 27 kDa protein 1	0.00364	0.31892
P00558	5230	PGK1	44.5/4.2	Phosphoglycerate kinase 1	0.50402	0.79456
Q01995	5876	TAGLN	22.5/4.3	Transgelin	−0.34128	−0.26434
P08631	3055	JTK9	59.5/6.8	Hemopoietic cell kinase	0.01446	0.02549
P09382	3956	LGALS1	16/5.5	Galectin-1, galactoside-binding, soluble, 1	0.026623	0.01123
Q92784	8110	DPF3	25.8/4.8	D4, zinc and double PHD fingers, family 3	0.094884	−0.03214
P54257	9001	HAP1	75.5/6.5	Huntington-associated protein 1	0.12364	−0.08108
P05783	3875	KRT18	48/5.3	Keratin 18	0.010218	0.60544
P05787	3856	KRT8	9.2/4.4	Keratin 8	0.041280	**0.84566**
P00738	3240	HP	55.2/6.2	Haptoglobin	0.044679	−0.14501
P09769	2268	FGR	59.5/5.2	Gardner–Rasheed feline sarcoma viral oncogene homolog	0.031264	−0.13789
P19838	4790	NFKB1	50.4/6.3	Nuclear factor of kappa light gene enhancer in B-cells 1	0.04467	−0.14501
P29034	6273	S100A2	10.9/4.6	S100 calcium-binding protein A2	**0.87964**	0.243214
Q13105	7709	ZBTB17	87.9/5.3	Zinc-finger and BTB domain containing 17	−0.17636	0.048108
Q00987	4193	MDM2	75.2/4.8	Transformed 3T3 cell double minute 2	−0.19782	−0.50544
P27816	4134	MAP4	111/5.4	Microtubule-associated protein 4	0.25872	−0.05356
P52732	3832	KIF11	119.2/6.2	Kinesin family member 11	−0.25778	−0.53444
P25205	4172	MCM3	90.9/5.5	Minichromosome maintenance deficient 3	0.25644	0.053666
P08631	3055	HCK	59.5/5.7	Hemopoietic cell kinase	0.65533	0.054501
P09237	4316	MMP7	22.6/5.8	Matrix metalloproteinase 7	0.234987	**0.876820**
P30305	994	CDC25B	64.9/4.5	Cell division cycle 25B	0.045116	0.283214
P50290	998	CDC42	21.3/6.1	Cell division cycle 42 (GTP-binding protein, 25 kDa)	−0.47636	0.088108
P61586	387	RHOA	19.8/6.9	Ras homolog gene family, member A	−0.49782	−0.00544
P63000	5879	RAC1	21.5/6.8	Ras-related C3 botulinum toxin substrate 1	−0.05583	−0.03566
P07437	203068	TUBB	49.6/6.5	Tubulin, beta polypeptide	0.255533	0.145010
P24864	898	CCNE1	47.1/4.3	Cyclin E1	−0.65116	0.232149
P04141	1437	CSF2	16.9/6.3	Colony stimulating factor 2 (granulocyte–macrophage)	−0.64636	−1.28108
P28072	5694	PSMB6	25.3/5.2	Proteasome (prosome, macropain) subunit, beta type, 6	−0.69782	−1.30544
P00352	216	ALDH1A1	54.7/4.3	Aldehyde dehydrogenase 1 family, member A1	−0.75872	0.03356
Q03013	2948	GTM4	25.3/5.0	Glutathione *S*-transferase M4	−0.78533	0.134501
P63241	1984	EIF5A	10/4.4	Eukaryotic translation initiation factor 5A	−0.97893	−1.44321
Q01469	2171	EFABP	18.0/4.2	Fatty acid-binding protein 5	0.25684	−0.36432

[a]Only the gene showing statistically significant increased protein expression with increases in both genomic copy number and transcript simultaneously will be considered as potential putative oncogene in lung adenocarcinoma cells. [b]*r*, Spearman correlation coefficients between proteins and genomic or mRNA values are based on all six cancer cell lines; bold indicates $P<0.05$, if $r>0.84000$. Mw, molecular weight; p*I*, isoelectric point.

immunohistochemical analysis), these findings are consistent with those from our cell lines, demonstrating again that genomic amplification and consequent increases in amounts of transcript may be, at least in part, driving the abundance of proteins in these lung tumors. The association between expression of these genes and that of Ki-67, a known indicator of poor prognosis in lung cancer (Martin *et al*., 2004), suggests that activation of these genes may be an indicator of tumor aggressiveness. These results also suggest that expression of EEF1A2 and KCIP-1 proteins in stage I non-small-cell lung cancer may be useful as a marker for distinguishing patients with relatively poor prognosis from those who might benefit from adjuvant treatment.

Discussion

Our current study illustrates the power of integrated functional genomic analyses for identifying putative oncogenes and for evaluating their potential clinical significance. Among the four identified oncogenes, three genes (*PRDX1*, *CALR*, and *KCIP-1*) have been implicated in lung tumorigenesis. PRDX1 is an antioxidant protein involved in regulating cell proliferation, differentiation, and apoptosis. Kim *et al*. (2003) found PRDX1 expression to be elevated in both lung cancer and adjacent normal lung tissue, suggesting that activation of *PRDX1* may enhance proliferation in lung cancer. *CALR* has a major role in Ca^{2+} binding and the

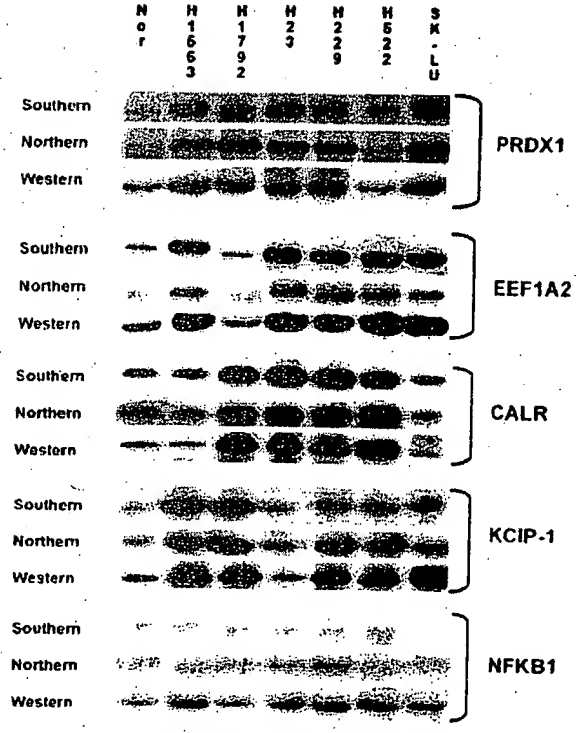

Figure 1 Confirmation by Southern, Northern, and Western blot analyses of increased DNA copies, transcript levels, and protein levels in the four genes identified in high-throughput analyses. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increased protein level did not correlate with genetic changes. The blotting results are consistent with the results from the CGH array, transcript array, and proteomic analyses. Nor, indicates normal bronchial epithelial cell line. All the experiments were repeated at least three times with each cell line. Means of normalized to β-actin signal intensities on Southern, Northern, and Western blots, along with 95% confidence intervals, were calculated (β-actin signals are not shown in the figure; two different normal bronchial epithelial cell lines were used in the confirmation and only one normal cell line is shown in the figure).

transcriptional regulation of other genes and was recently found to be overexpressed in 73% of 40 lung adenocarcinomas (Oates and Edwards, 2000). *KCIP-1* belongs to the 14-3-3 family, which participates via the MAPK and Wnt signaling pathways in the regulation of many cellular processes including cell proliferation and differentiation as well as tumorigenesis (Thomas *et al.*, 2005). *KCIP-1* was recently found to be expressed in all 12 lung tumors tested in a single-institution study (Qi *et al.*, 2005). Interestingly, *EEF1A2* was originally considered a putative oncogene in ovarian cancer on the basis of its being amplified in 25% and overexpressed in 30% of the same set of ovarian tumors (Anand *et al.*, 2002); functional analyses have established its oncogenic role in cellular transformation (Lee, 2003). Our discovery that *EEF1A2* may be a putative oncogene in lung adenocarcinoma demonstrates the power of our functional genomic strategy for rapidly identifying potential oncogenes.

Although the main focus of this study was to specifically identify putative oncogenes, it should be noted that 90.7% of the genes showing high protein expression did not show corresponding increases in both DNA copy number and transcript, a finding consistent with that of others that transcriptional, translational, and post-translational regulatory mechanisms can greatly influence the abundance of protein in lung tumorigenesis (Chen *et al.*, 2002). For example, *NFKB1* is a critical arbiter of immune responses, cell survival, and transformation and is often activated in several types of tumors (Chen *et al.*, 2002). Deregulation of *NFKB1* is thought to be modulated through phosphorylation of Ser337 by protein kinase A (Chen *et al.*, 2002). In our study, 68.8% of the genes showing over-representation in the genome did not show elevated transcript levels, implying that at least some of these genes are 'passenger' genes that are concurrently amplified because of their location with respect to amplicons but lack biological relevance in terms of the development of lung adenocarcinoma.

Although the potential oncogenes we identified here are likely to be important, certainly other oncogenes could be involved in the development of lung adenocarcinoma. The oligo microarray we used consists of 22 000 probes, which represent only about 60% of the human genome. Moreover, each probe was designed for the 3′ region of expressed sequence tags of the selected genes. Also, our results were initially derived from cancer cell lines, although the findings were later confirmed in human tissue samples. Our ongoing study using microarrays with information on more genes and the development of high-resolution proteomic analyses for use with larger numbers of specimens will allow more comprehensive analyses of the molecular consequences of gene amplifications. Such expanded analyses will very likely lead to the identification of additional oncogenes.

Some of the results of our current study were comparable to those of other studies of lung cancer. For example, genomic copy number and protein levels of KCIP-1 were previously found to be amplified and overexpressed in primary lung cancers by cDNA clone-based CGH array analysis (Jiang *et al.*, 2004) and proteomic analysis (Chen *et al.*, 2002), respectively. Our functional genomic approach, which integrates simultaneous CGH, transcript microarrys, proteomic analyses, and siRNA, allows us not only to quickly identify potential oncogenes but also to explore their significance as diagnostic and therapeutic targets in tumor progression – more than could be achieved by any technique alone.

Genes identified in this way may serve as promising targets for diagnosis and therapy in lung adenocarcinoma. Further research on the clinical implications of such genes is needed; experiments now underway in our laboratory include overexpression of the genes in normal cells, disruption of the function of these genes in cancer cells, and investigation of how interactions among these genes (or interactions with other known oncogenes) may mediate the expression of the transformed phenotype.

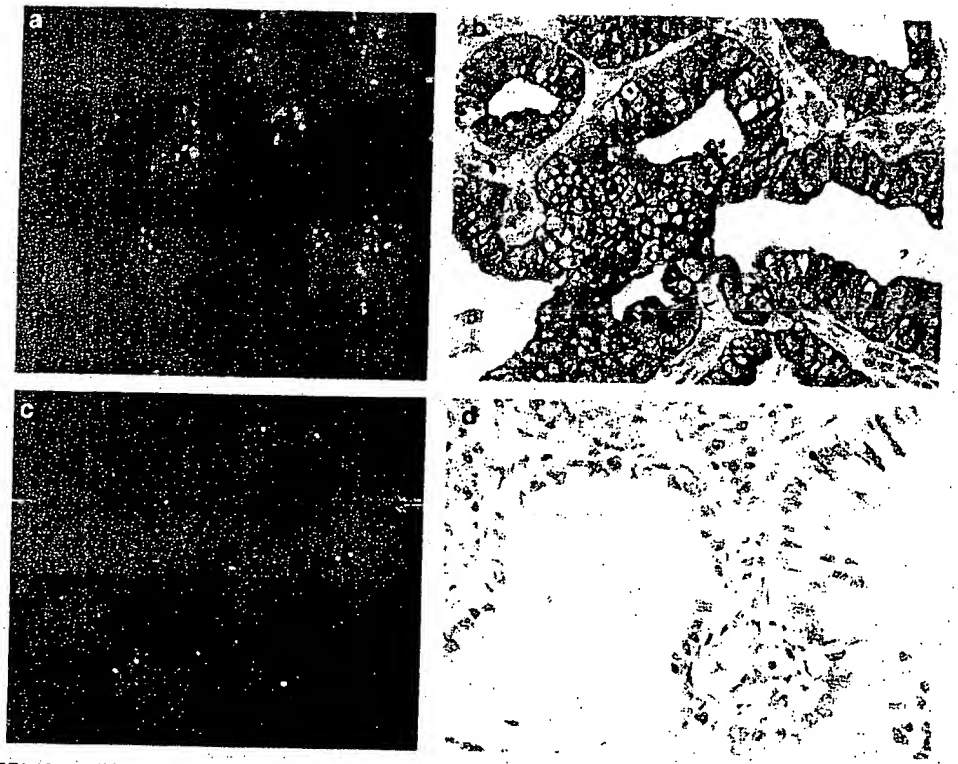

Figure 2 *EEF1A2* amplification is associated with high EEF1A2 protein expression in lung adenocarcinomas. (**a**) Cells from a lung adenocarcinoma sample in which *EEF1A2* is amplified show more green signals (*EEF1A2*) than red signals (chromosome 20 centromeric probe) (original magnification, × 400). (**b**) Immunohistochemical staining of cells from the same tissue sample as in panel a shows strong EEF1A2 staining in the cytoplasm. (**c**) A lung adenocarcinoma sample with two copies of *EEF1A2* and chromosome 20 centromeric probe, indicating no *EEF1A2* amplification (original magnification, × 400). (**d**) Immunohistochemical staining of cells from the same tissue sample as in panel c shows negative staining for EEF1A2.

Materials and methods

Cell lines

Six human lung adenocarcinoma cell lines (H23, H229, H1792, SK-LU-1, H522, and H1563) were obtained from the American Type Culture Collection (Manassas, VA, USA). Two normal bronchial epithelial cell lines were obtained from Clontech (Palo Alto, CA, USA). Genomic DNA, mRNA, and protein were derived from a single harvest of these cells.

DNA and RNA profiles by microarray analysis

Genomic DNA labeling and hybridization were performed as described previously (Barrett *et al*., 2004) with Agilent's Human 1A Oligo Microarray (V2) (Agilent Technologies, Palo Alto, CA, USA), which contains 22 000 unique 60-mer oligos. Details of the protocol for analysing transcripts are available at http://www.chem.agilent.com. Map positions for arrayed genes were assigned by identifying the DNA sequence represented in the UniGene cluster and matching it with the Golden Path genome assembly (http://genome.ucsc.edu/; Mat 7, 2004 Freeze). Microarray images of DNA copy number and expression were analysed by using AgilentCGH Analytics and Feature Extraction software. DNA copy number profiles that deviated significantly from background signal ratios (measured from normal control cell hybridization, as described elsewhere; Barrett *et al*., 2004) were interpreted as evidence of true differences in DNA copy number. The criteria for defining genomic over-representation and amplicons are described elsewhere (Hyman *et al*., 2002); details are given in the Supplementary Information. An increase in mRNA level was defined as a twofold increase in signal ratio relative to that of the control ($\log_2 > 1$).

Quantitative two-dimensional PAGE and mass spectrometry

Analysis of proteins by two-dimensional PAGE and their identification by mass spectrometry were performed as previously described (Shen *et al*., 2004). Briefly, protein pellets were solubilized in rehydration buffer, after which the first-dimension isoelectric focusing was carried out with a Protean IEF Cell (Bio-Rad Laboratories) and the second-dimension separation was carried out with Bio-Rad's Ready Gel Precast Gels and the Bio-Rad Criterion Cell apparatus. Protein spots were visualized by silver-based staining, and all gels were assessed with Bio-Rad's PDQuest 2D gel image analysis software. Selected spots were subjected to in-gel tryptic digestion and analysed on a Voyager-DE PRO matrix-assisted laser desorption ionization/time-of-flight mass spectrometer (Applied Biosystems, Foster City, CA, USA). The mass list of the 20 most intense monoisotopic peaks for each sample was entered in the MS-Fit search program (v3.2.1) (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm) and searched in the National Center for Biotechnology Information protein database.

Southern, Northern, and Western blot analyses

Southern, Northern, and Western blot hybridizations were performed according to standard protocols. cDNA clones for the tested genes were purchased from Invitrogen (Carlsbad,

CA, USA) and prepared as probes for the blot hybridizations. Antibodies used were obtained as follows: PRDX1, CALR, NFKB1, KCIP-1, and β-actin from Santa Cruz Biotechnology (Santa Cruz, CA, USA); and EEF1A2 from Upstate Biotechnology (Waltham, MA, USA).

Fluorescence in situ *hybridization and immunohistochemical analyses of lung tissue microarrays*

Fluorescence *in situ* hybridizations and immunohistochemical analyses of KCIP-1 and EEF1A2 were carried out as described elsewhere (Jiang *et al*., 2002; Wang *et al*., 2005) with Lung Tissue Microarrays (Ambion, Austin, TX, USA) and 11 homemade microarray blocks containing tissue samples from 113 patients with pathologic stage I non-small-cell lung cancer (Wang *et al*., 2005). DNA probes specific for *KCIP-1* and *EEF1A2* were obtained by screening a Human BAC Clone library (Invitrogen) by polymerase chain reaction as described previously (Jiang *et al*., 2002). The antibodies used for the immunohistochemical analyses were the same as those used for the Western blotting. Cell proliferation of the lung tissues was assessed with a Ki-67 monoclonal antibody from Santa Cruz Biotechnology. Definitions of the cutoff value for a positive result of each antibody are shown in Supplementary Information.

siRNA transfection, cellular proliferation assay, and apoptosis analysis

Transfections were carried out by using siPORT Lipid Transfection Agent (Ambion) with siRNAs targeting KCIP-1 or EEF1A2 or with a scrambled siRNA duplex (siControl) (Dharmacon Inc., Lafayette, CO, USA), with PBS used as a negative control (Jiang *et al*., 2002). Cells were fixed 24, 48, or 96 h later and subjected to further tests. All siRNAs were prepared by using a transcription-based method with Silencer siRNA according to the manufacturer's instructions (Ambion). Sequences of the individual siRNAs are listed in Supplementary Table 4S. Inhibition of cell growth by the siRNAs was determined by MTT staining, and cell growth rate was plotted against the percentage of viable cells in the saline-treated controls (a value arbitrarily set at 100%) (Jiang *et al*., 2002). Apoptosis was analysed by fluorescence cell cycle analysis of terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling with FITC-labeled dUTP (Boehringer Mannheim Biochemicals, Mannheim, Germany) (Jiang *et al*., 2005).

Statistical analyses

Relationships between gene copy number and mRNA level were examined as described elsewhere (Hyman *et al*., 2002, Supplementary Information). Correlations between protein abundance and DNA copy number and mRNA expression of the corresponding genes were evaluated with the Spearman correlation coefficient. Fisher's exact test and χ^2-tests were used to analyse associations between amplification and expression of the candidate genes with various histopathologic variables of the samples in the tissue microarrays. Univariate and multivariate analyses were carried out with Cox's proportional hazards model to determine which independent factors might have a joint significant influence on survival. A *P*-value $\leqslant 0.05$ was considered statistically significant; all statistical tests were based on a two-sided significance level.

Acknowledgements

This work was supported by National Institutes of Health Grant CA113707-01, an institutional research grant from The University of Texas MD Anderson Cancer Center, a Developmental Project/Career Development Award from The University of Texas Specialized Programs of Research Excellence in Lung Cancer P50 CA70907, and an M Keck Center for Cancer Gene Therapy Award (all to FJ). We thank Christine F Wogan of the Department of Scientific Publications for editorial review of this manuscript.

References

Anand N, Murthy S, Amann G, Wernick M, Porter LA, Cukier IH *et al*. (2002). *Nat Genet* **31**: 301–305.

Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R *et al*. (2004). *Proc Natl Acad Sci USA* **101**: 17765–17770.

Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardia SL *et al*. (2002). *Mol Cell Proteomics* **1**: 304–313.

Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E *et al*. (2002). *Cancer Res* **62**: 6240–6245.

Jiang F, Caraway NP, Li RY, Katz RL. (2005). *Oncogene* **24**: 3409–3418.

Jiang F, Lin F, Price R, Gu J, Medeiros LJ, Zhang HZ *et al*. (2002). *J Mol Diagn* **4**: 144–149.

Jiang F, Yin Z, Caraway NP, Li R, Katz RL. (2004). *Neoplasia (New York)* **6**: 623–635.

Kim HJ, Chae HZ, Kim YJ, Kim YH, Hwangs TS, Park EM *et al*. (2003). *Cell Biol Toxicol* **19**: 285–298.

Lee JM. (2003). *Reprod Biol Endocrinol* **1**: 69–73.

Luk C, Tsao MS, Bayani J, Shepherd F, Squire JA. (2001). *Cancer Genet Cytogenet* **125**: 87–99.

Martin B, Paesmans M, Mascaux C, Berghmans T, Lothaire P, Meert AP *et al*. (2004). *Br J Cancer* **91**: 2018–2225.

Oates J, Edwards C. (2000). *Histopathology* **36**: 341–347.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE *et al*. (2002). *Proc Natl Acad Sci USA* **99**: 12963–12968.

Qi W, Liu X, Qiao D, Martinez JD. (2005). *Int J Cancer* **113**: 359–363.

Ross JS, Fletcher JA. (1999). *Semin Cancer Biol* **9**: 125–138.

Shen J, Person MD, Zhu J, Abbruzzese JL, Li D. (2004). *Cancer Res* **64**: 9018–9026.

Thomas D, Guthridge M, Woodcock J, Lopez A. (2005). *Curr Top Dev Biol* **67**: 285–303.

Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y *et al*. (2005). *Proc Natl Acad Sci USA* **102**: 9625–9630.

Wang H, Zhang Z, Li R, Ang KK, Zhang H, Caraway NP *et al*. (2005). *Int J Cancer* **116**: 285–290.

Supplementary Information accompanies the paper on Oncogene website (http://www.nature.com/onc)